DOCUMENT RESUME

ED 421 197 JC 980 346

AUTHOR	Arnold, Carolyn L.
TITLE	Using National Data Sets To Create Comparable National Statistics for the Student Characteristics and Outcomes in Community Colleges.
INSTITUTION	Chabot Coll., Hayward, CA.
SPONS AGENCY	National Center for Education Statistics (ED), Washington, DC.; National Science Foundation, Arlington, VA.; Association for Institutional Research.
PUB DATE	1997-10-00
NOTE	83p.
PUB TYPE	Reports - Evaluative (142)
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	Cluster Grouping; College Outcomes Assessment; *Community Colleges; *Comparative Analysis; Data Collection; *Institutional Research; *Peer Institutions; Statistical Analysis; Tables (Data); Two Year Colleges
IDENTIFIERS	Beginning Postsecondary Students Long Study; Chabot College CA; Integrated Postsecondary Education Data System; National Center for Education Statistics

ABSTRACT

Addressed to institutional researchers, this report from California's Chabot College presents information on National Center for Education Statistics (NCES) data sets. Included is a discussion on how these data sets can be used to create peer groups of U.S. colleges, and to produce statistics on major student variables for each of these groups. Ideas are presented on how to obtain and work with NCES data sets, and what resources are needed in terms of time, hardware, software, supplies, and funds. Three major chapters review the usefulness of two data sets: the Integrated Postsecondary Education Data System (IPEDS) and the Beginning Postsecondary Students Longitudinal Study (BPS). The conclusion summarizes the value and accessibility of these data sets, and helps institutional researchers evaluate whether they have the inclination or resources to work with either of the sets, or if published materials are better alternatives. The report contains 21 tables. The appendices contain lists of IPEDS and BPS variables used, supporting tables, and additional sources of help in using the NCES data sets. (AS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 421 197

Chabot College

Office of Institutional Research

Using National Data Sets to Create Comparable National Statistics for the Student Characteristics and Outcomes in Community Colleges

October, 1997

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

C. L. Arnold

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Carolyn L. Arnold, Ph.D.
Institutional Researcher

Chabot College
25555 Hesperian Blvd.
Hayward, California 94545
carnold@clpccd.cc.ca.us
(510) 786-6965

JC 980 346

Chabot College

Office of Institutional Research

Using National Data Sets to Create Comparable National Statistics for the Student Characteristics and Outcomes in Community Colleges

October, 1997

Carolyn L. Arnold, Ph.D.
Institutional Researcher

Chabot College
25555 Hesperian Blvd.
Hayward, California 94545
carnold@clpccd.cc.ca.us
(510) 786-6965

Acknowledgments

This project was supported by the Association for Institutional Research (AIR) Research Grant Program, *Improving Institutional Research in Postsecondary Educational Institutions,* with funds provided by the National Center for Education Statistics (NCES) and the National Science Foundation (NSF), under Association for Institutional Research Grant No. 207.

The author wishes to thank the many people at AIR, NCES, and Chabot College who helped make this research grant and report possible. First, I want to thank AIR, NCES, and NSF for setting up the research grant program, which allows institutional researchers the opportunity to become experienced with the national data sets and thus have a much more national perspective as well an appreciation of the research conducted at NCES and NSF. My appreciation goes to Terry Russell of AIR, who worked with me to interpret the reviewer's comments and make the grant feasible, and to Ann MacMillan of AIR, who provided cheery logistical advice throughout the year. Both Terry and Ann made the process of refining and working on the grant very enjoyable and trouble-free, and they made those of us who received these grants feel very special at the 1997 AIR Forum in Orlando.

At NCES, I can't thank Aurora D'Amico enough for doing her job so well—providing the public with information on and access to NCES data sets. She was always there to help when the official channels were unclear, and she stayed in contact with me about each problem until it was solved. She became my first point of contact and personal shopper, advocate, and facilitator for any data sets, manuals, answers about data set quirks, or statistical advice that I needed. The people she put me in touch with—Bill Freund, Andrew Malizio, Larry Bobbitt, and Sam Peng—were very generous with their thoughtful advice and answers to my many e-mails. Dennis Carroll and Carl Schmidt, my previous contacts at NCES, provided me with timely information about data, software, and statistics as well. I appreciated all their advice, and I take full responsibility for not always following it. Also doing their jobs well were Alan Moorehead, former Data Security Officer, and Tracy Ferbisch at the NEDRC who responded to my (and Aurora's) requests.

At Chabot College, Cieny Carney and Precious Chambers of the Grants Development Office gave me both moral and logistical support in preparing the grant. Their support and celebration of campus grant writers, no matter what the outcome, makes even the grant-writing process rewarding. Mike Calegari of the Chabot Business Office, championed the grant through the many financial and bureaucratic hoops needed to set up a working grant account. And finally, I could not have accomplished all that I did on this grant during the academic year without the gracious and steadfast support of my boss, Donna Marie Ferro, Dean of Matriculation and Academic Standards. She both encouraged me to apply for the grant, and understood when I had to work on it.

Disclaimer

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the Association for Institutional Research, the National Center for Education Statistics, or the National Science Foundation.

Table of Contents

page number

Acknowledgments and Disclaimer ii
Table of Contents iii
List of Text Tables v
List of Appendix Tables vii

I. Introduction
- Purpose of report 1
- Overview of data sets used 1
- Report organization 2

II. Usefulness: Identifying peer variables in IPEDS
- Peer group literature 3
- Peer group variables in IPEDS 4
- The search for peer group variables for the community colleges 6
  - IPEDS variables in BPS 6
  - IPEDS variables in IPEDS only 9
- Conclusion 14

III. Usefulness: Finding IR student characteristics and outcome variables in BPS
- Common student variables used by community colleges 15
- IR student characteristics and outcomes variables in BPS 16
- Student characteristics and outcomes by sector and peer groups 17
  - Differences by sector 28
  - Differences by diversity 28
  - Differences by enrollment 28
  - Differences by percentage of full-time students 28
- Overall findings 29

IV. Accessibility: How to obtain and work with NCES datasets
- Types of datasets available 31
  - IPEDS public-use files 31
  - NPSAS and BPS 31
  - NPSAS and BPS public-use files 32
  - NPSAS and BPS restricted-use files 32
- Preliminary considerations: hardware/software, time, funds 33
  - Hardware/software requirements 33
  - Time requirements 34
  - Funds/supplies/computer support requirements 35
- How to obtain data and reports: logistics 36
  - IPEDS public-use files 36
  - NPSAS and BPS public-use files 36
  - NPSAS and BPS restricted-use files 36
- Challenges when working with restricted-use data sets 37
  - Choosing variables: IPEDS 37
  - Choosing variables: NPSAS/BPS 38
  - Choosing and normalizing weights 39
  - Estimating standard errors 40
  - Testing for differences 42
- Summary of steps for obtaining and working with restricted-use files 43

Table of Contents (continued)

V. Conclusion

Usefulness and accessibility of NCES data sets 45
Who should use the restricted-use and public-use data sets? 46
What are the alternatives? 46

VI. Appendices

A. Variable lists

List of IPEDS variables examined A-1
IPEDS Analysis File layout A-2
List of BPS variables examined A-3

B. Supporting tables for Chapter II

Standard errors tables B-1
IPEDS percentage distribution of percent women and AA/AS degrees B-3
IPEDS peer variable crosstabulations B-4

C. Supporting tables for Chapter III*

Examples: estimating standard errors with design effects; calculating t-tests.. C-1
Standard errors tables C-2

D. Contact names and addresses for ordering NCES data sets and publications..... D-1

***Note on numbering of Appendix C tables:**
Appendix C contains tables C1 and tables C12a through C16b. Table C1 provides examples of calculating standard errors using design effects and testing for differences using standard errors. Tables C12a through C16b contain the standard errors for the estimates in text tables 12a through 16b in Chapter III. In order to make it easier to locate the appropriate standard errors, these tables in Appendix C are numbered based on the corresponding text table, i.e., appendix table C12a contains the standard errors for text table 12a. Therefore, tables C2 through C11 do not exist.

List of Text Tables

page number

Table 1. Variables in IPEDS with potential as peer group variables 5

Table 2. Percentage distribution of race-ethnicity of 1989-90 beginning postsecondary students, by selected institutional sectors 7

Table 3. Percentage distribution of race-ethnicity of 1989-90 beginning postsecondary students in public 2-year colleges, by enrollment size 7

Table 4. Percentage distribution of race-ethnicity of 1989-90 beginning postsecondary students in public 2-year colleges, by selected regions 8

Table 5. Percentage distribution of race-ethnicity of 1989-90 beginning postsecondary students in public 2-year colleges, by selected states 8

Table 6. Percentage distribution of selected peer group variables in 1990 IPEDS population of public two-year colleges and BPS sample of public two-year students 10

Table 7. Percentage distribution of race-ethnicity of 1989-90 beginning postsecondary students in public 2-year colleges, by percentage of students of color in student body in three categories 12

Table 8. Percentage distribution of race-ethnicity of 1989-90 beginning postsecondary students in public 2-year colleges, by percentage of students of color in student body in two categories 12

Table 9. Percentage distribution of race-ethnicity of 1989-90 beginning postsecondary students in public 2-year colleges, by enrollment size of college 13

Table 10. Percentage distribution of race-ethnicity of 1989-90 beginning postsecondary students in public 2-year colleges, by percentage of full-time students in student body... 13

Table 11. Types of student characteristics and outcomes variables available in BPS 16

Table 12a. Percentage distribution of gender, race-ethnicity, citizenship, age, and religion of beginning postsecondary students, by sector and by percentage of students of color within public two-year colleges, 1989-90 18

Table 12b. Percentage distribution of gender, race-ethnicity, citizenship, age, and religion of beginning postsecondary students in public two-year colleges, by enrollment size and percentage of full-time students: 1989-90 19

Table 13a. Percentage distribution of SES, parent's education, high school diploma and years since high school, marital status and number of children of beginning postsecondary students, by sector and by percentage of students of color within public two-year colleges, 1989-90 20

List of Text Tables (continued)

page number

Table 13b. Percentage distribution of SES, parent's education, high school diploma and years since high school, marital status and number of children of beginning postsecondary students in public two-year colleges, by enrollment size and percentage of full-time students: 1989-90 21

Table 14a. Percentage distribution of distance from home, residence, financial dependency, college attendance, and major of beginning postsecondary students, by sector and by percentage of students of color within public two-year colleges, 1989-90 22

Table 14b. Percentage distribution of distance from home, residence, financial dependency, college attendance, and major of beginning postsecondary students in public two-year colleges, by enrollment size and percentage of full-time students: 1989-90 23

Table 15a. Percentage distribution of work hours, current occupation, and future occupation goal of beginning postsecondary students, by sector and by percentage of students of color within public two-year colleges, 1989-90 24

Table 15b. Percentage distribution of work hours, current occupation, and future occupation goal of beginning postsecondary students in public two-year colleges, by enrollment size and percentage of full-time students: 1989-90 25

Table 16a. Percentage distribution of degree goal, educational aspirations, persistence and attainment, and degree progression of beginning postsecondary students, by sector and by percentage of students of color within public two-year colleges, 1989-90 26

Table 16b. Percentage distribution of degree goal, educational aspirations, persistence and attainment, and degree progression of beginning postsecondary students in public two-year colleges, by enrollment size and percentage of full-time students: 1989-90 27

List of Appendix Tables

page number

Appendix A. Variable lists

Table A1. Variables in IPEDS with potential for use in identifying peer groups A-1

Table A2. Record layout for IPEDS 1990 Analysis File with shorter records A-2

Table A3. Student characteristics and outcomes variables examined in BPS A-3

Appendix B. Supporting tables for Chapter II

Table B1. Standard errors for percentage distribution of race-ethnicity of 1989-90 beginning postsecondary students, by selected sectors B-1

Table B2. Standard errors for percentage distribution of race-ethnicity of 1989-90 beginning postsecondary students in public 2-year colleges, by enrollment size, selected regions, and selected states B-2

Table B3. Percentage distribution of public two-year colleges in IPEDS 1990, by selected student demographic and outcome measures B-3

Table B4. Crosstabulations and chi-squared values of potential peer group variables for public two-year colleges in IPEDS 1990 ..... B-4

Appendix C. Supporting tables for Chapter III

Table C1. Examples of using DAS standard errors to estimate the DAS design effect and using this design effect to estimate the standard errors for non-DAS variables and perform t-tests on the differences between the percentages of selected student characteristics and outcomes of beginning postsecondary students in public two-year colleges C-1

Table C12a. Standard errors for percentage distribution of gender, race-ethnicity, citizenship, and age, religion of beginning postsecondary students, by sector and by percentage of students of color within public two-year colleges, 1989-90 ..... C-2

Table C12b. Standard errors for percentage distribution of gender, race-ethnicity, citizenship, age, and religion of beginning postsecondary students in public two-year colleges, by enrollment size and percentage of full-time students: 1989-90 ..... C-3

Table C13a. Standard errors for percentage distribution of SES, parent's education, high school diploma and years since high school, marital status and number of children of beginning postsecondary students, by sector and by percentage of students of color within public two-year colleges, 1989-90 ..... C-4

Table C13b. Standard errors for percentage distribution of SES, parent's education, high school diploma and years since high school, marital status and number of children of beginning postsecondary students in public two-year colleges, by enrollment size and percentage of full-time students: 1989-90 ..... C-5

List of Appendix Tables (continued)

page number

Table C14a. Standard errors for percentage distribution of distance from home, residence, financial dependency, college attendance, and major of beginning postsecondary students, by sector and by percentage of students of color within public two-year colleges, 1989-90 ..... C-6

Table C14b. Standard errors for percentage distribution of distance from home, residence, financial dependency, college attendance, and major of beginning postsecondary students in public two-year colleges, by enrollment size and percentage of full-time students: 1989-90 ..... C-7

Table C15a. Standard errors for percentage distribution of work hours, current occupation, and future occupation goal of beginning postsecondary students, by sector and by percentage of students of color within public two-year colleges, 1989-90 ..... C-8

Table C15b. Standard errors for percentage distribution of work hours, current occupation, and future occupation goal of beginning postsecondary students in public two-year colleges, by enrollment size and percentage of full-time students: 1989-90 ..... C-9

Table C16a. Standard errors for percentage distribution of degree goal, educational aspirations, persistence and attainment, and degree progression of beginning postsecondary students, by sector and by percentage of students of color within public two-year colleges, 1989-90 ..... C-10

Table C16b. Standard errors for percentage distribution of degree goal, educational aspirations, persistence and attainment, and degree progression of beginning postsecondary students in public two-year colleges, by enrollment size and percentage of full-time students: 1989-90 ..... C-11

Chapter 1
Introduction

I. Introduction

Purpose of report

This report demonstrates how institutional researchers in colleges and universities can use U.S. Department of Education national data sets to create groups of comparable 'peer' colleges across the U.S. and produce statistics on major institutional research (IR) student variables for these peer groups. Comparative information is becoming mandatory for institutional research reports.[1] National peer group statistics address the need of researchers to offer a more meaningful national comparison for their college's student characteristics and outcomes than a national average. This research project was funded by the Association for Institutional Research (AIR), National Center for Education Statistics (NCES), and the National Science Foundation (NSF) under the AIR grant program called *Improving Institutional Research in Postsecondary Educational Institutions*, and is addressed to institutional researchers.

The major questions answered by this report are how useful and how accessible these NCES data sets are to institutional researchers, particularly those in community colleges where student populations vary substantially and the research staff and funding are low. To be useful, researchers must be able to produce comparable statistics on relevant variables from these data sets. This report will discuss whether the data sets contain variables that can be used to create peer groups of colleges and that reflect common IR student characteristics and outcomes used in community colleges. To be accessible, the resources and skills needed to work with these data sets must be within the range of the average institutional researcher in a small office. This report will discuss how to obtain and work with the data sets, and what resources are needed in terms of time, hardware, software, supplies, and funds. It will also help researchers evaluate whether the effort is worth it in their situation, or whether the published materials are better alternatives.

Overview of data sets used

Two NCES data sets were used to find variables for peer groups and IR variables—the Integrated Post-secondary Education Data System (IPEDS) and the Beginning Postsecondary Students Longitudinal Study (BPS).[2] For institutional-level variables that could define peer groups, IPEDS was used. This population data set contains information on all postsecondary

---

[1] Joseph L. Marks, "Toward a New Breed of Fact Book," *New Directions for Institutional Research*, No 91, Fall 1996.

[2] These were the most appropriate NCES data sets for this project. However, NCES conducts several other studies that are relevant for postsecondary education research. See Appendix D for information sources on all NCES data sets.

institutions in the U.S. and is collected each year. Data is collected from colleges on fall enrollments, completions, institutional characteristics, faculty salaries, and finance, and is stored in five files of the same names. The 1990 files were used for this report because BPS students started in 1989-90, and the 1990 IPEDS files contain both 1989 and 1990 information about their institutions .

For student-level variables that could provide student characteristics and outcomes, BPS was used. This data set is a longitudinal study of a national sample of students in their first year of college in 1989-90 in all types of postsecondary institutions. These students were followed through 1994, whether or not they stayed in college. The students in this study were obtained from a 1989-90 cross-sectional study, the National Postsecondary Student Aid Study (NPSAS), that surveyed a national sample of students at all levels in all types of postsecondary institutions in 1989-90. NPSAS was conducted in 1989-90 and again in 1992-93, and its major purpose was to determine how students funded their college education. BPS took the students in the 1989-90 NPSAS who were in their first year of college that year and followed them. BPS is an institutional researcher's dream data set—it contains a wealth of variables on postsecondary student characteristics and outcomes. With its origin in NPSAS, BPS is especially strong on the details of financial status, financial aid, and family data, but it also includes a range of other student characteristics as well as many longitudinal student outcomes such as persistence and transfer. In addition, BPS also contains some institutional variables from IPEDS.

Report organization

This report contains three major chapters. Chapter II addresses the usefulness of the IPEDS data set for creating peer groups among community colleges by exploring the IPEDS files for appropriate variables. The process of identifying peer groups is illustrated with the search for peer group variables that are meaningful to California community colleges. Chapter III reviews the usefulness of the BPS data set for identifying key institutional research variables on student characteristics and outcomes by presenting data on many of these variables, by sector and by peer groups within the community college sector. Chapter IV addresses the accessibility of the data sets by describing the process of obtaining and using both the public-use and restricted-use data sets. The conclusion in Chapter V summarizes the conclusions about usefulness and accessibility and helps institutional researchers evaluate whether they have the inclination or resources to work with either of these types of data sets. The appendices contain lists of variables examined or used from each data set, some supporting tables, tables of standard errors, examples of estimating complex standard errors and performing t-tests, and lists of NCES names and addresses for ordering and obtaining help with the data sets.

Chapter II

Usefulness
Identifying peer variables in IPEDS

II. Usefulness
Identifying peer variables in IPEDS

Peer group literature

Peer groups of postsecondary institutions are usually chosen in order to compare and evaluate institutional characteristics such as budget levels, faculty salaries, or program offerings to those of similar colleges.[3] The purpose of these comparisons is to determine whether the original college is in the normal range of these characteristics, or is higher or lower than its peers. In order to make these comparisons as valid as possible, the peer group must be chosen very carefully so it reflects the uniqueness of that college.[4] To accomplish this, as many quantitative and qualitative variables as possible are taken into account when defining the college and its peer group. In order to make distinctions as fine as possible, variables are sometimes evaluated with complex statistical methodologies, such as factor and cluster analysis.[5] Although the final analysis may result in only a few peer institutions, potential peers can be drawn from the entire population of postsecondary educational institutions in the U.S. Thus, colleges often start with the IPEDS data set as a source of those institutional-level variables.[6] They often add other quantitative variables from other sources such as the U.S. Census or local population and geographic data. Then the final group of peer institutions is usually selected or approved by high-level administrators after more qualitative

---

[3]Deborah J. Teeter and Paul T. Brinkman, "Peer Institutions" in M. A. Whiteley, J.D. Porter, and R. H. Fenske, eds., *The Primer for Institutional Research* (Association for Institutional Research, Resources for Institutional Research, Number Seven, Tallahassee, Florida, 1992). For examples see Nancy Ellen Soteriou, "Peer Institution Profile Report, El Paso Community College" (El Paso Community College, El Paso, Texas, 1994); Nathan Dickmeyer and Bradley Meeker, "Comparative Financial Statistics for Public Two-year Colleges: FY 1993 Peer Group Sample" (National Association of College and University Business Officers, Washington, D.C. 1994); State University of New York, "Central Administration Costs: Report 92-S-104" (New York State Office of the Comptroller, Albany, NY, 1993); "Tuition and Fee Rates, 1992-93. A National Comparison" (Washington State Higher Education Coordinating Board, Olympia, 1993); Craig A. Clagett, "A Community College Responds to a State Budget Crisis: An Evaluation of the Financial Plan for Prince George's Community College" (Office of Institutional Research and Analysis, Prince George's Community College, Largo, MD, 1992); "Accountability and Productivity: Report for the Illinois Community College System" (Illinois Community College Board, Springfield, 1992); "John C. Sutusky, "An Analysis of Tuition and Required Fees: South Carolina Public Colleges and Universities and Peer Institutions" (South Carolina Commission on Higher Education, Columbia, 1992); Richard M. Summerville and Dennis R. Ridley, "Toward Establishing Salary Benchmarks for College and University Administrators" (ERIC HE023820, 1990).

[4]Ryan Cherland, Principal Analyst in the Office of Institutional Research and Planning at the University of Kansas described a very careful process that whittled 45 possible peer institutions for Kansas State and University of Kansas to 14 for each college (E-mail correspondence, August 20, 1996). Ingram describes this process more formally in John A. Ingram, "Using IPEDS Data for Selecting Peer Institutions" (Paper presented at AIR Annual Forum, Boston, Mass, May, 1995)

[5]See Teeter and Brinkman, op. cit., Ingram, op. cit. and Carolyn L. Della Mea, "A comparison of two procedures for peer group assignment of institutions of higher education," (Unpublished dissertation, Virginia Polytechnic Institute and State University, Blacksburg, VA, 1989).

[6]Ingram, op. cit.

criteria, such as the mission and vision of the colleges, has been used to further refine the peer group. At that point, the colleges not selected for the peer group are of no concern.

In this study, the purpose of choosing peer groups in this study was also to provide comparison groups for evaluating institutional-level student characteristics. However, in contrast to the creation of one peer group for a particular college, the goal was to categorize all the public two-year community colleges into several peer groups that would be meaningful to the entire range of California community colleges. In this way, it is closer to a 'classification-based group' than a true peer group.[7] In addition, the variables used to create peer groups had to be easily accessible by institutional researchers with limited time and resources. Therefore, data that were readily available and relevant to all colleges had to be used, and qualitative information could not be considered.

The process of identifying these peer groups was less complex than the usual peer group analysis because the student characteristics to be estimated were obtained from a national sample of postsecondary institutions and students in BPS rather than from the entire population of institutions in IPEDS. Although IPEDS contains variables that can be used for peer group formations for all colleges in the U.S., the peer group analysis of the student characteristics in this study was limited to the student data from the sample of colleges that were also in BPS. Consequently, small sample sizes of colleges and students in BPS, especially in any one sector such as community colleges, necessitated the formation of only two or at the most three peer groups. With more than three groups, large margins of error for the estimates of the student demographics and outcomes would make any differences between the peer groups meaningless. Thus, although several variables in IPEDS were found to be salient for California community colleges and could theoretically be combined to form peer groups, only one variable at a time could realistically be used to differentiate two or three peer groups.

Peer group variables in IPEDS

IPEDS files contain the basic quantitative descriptors of postsecondary institutions. IPEDS consists of five files with information on fall enrollments, completions, institutional characteristics, faculty salaries, and finance. In addition, NCES has created a sixth file, called the Analysis file, which contains the most commonly used variables from each file.[8] Table 1 lists the IPEDS variables that were either found in the peer group literature, suggested by NCES staff, or had

---

[7]Based on classifications in Teeter and Brinkman, op. cit. However, the broader term 'peer group' will be used throughout this report.

[8]For a discussion of some difficulties in using the Analysis file, see the section on time requirements for public-use files in Chapter IV.

potential for use in identifying peer groups among community colleges.[9] These variables are listed based on whether they are included in the BPS file or are found only in IPEDS files, and whether they were tested in this study or not.

Table 1. Variables in IPEDS with potential as peer group variables

IPEDS variables in BPS

Tested in this study
Institutional sector (level and control)
Enrollment size
Region
State

Other
Type of calendar system
Admissions requirements (HS diploma, TOEFL, HS class standing, test scores, SAS, ACT, other tests, residency, ability to benefit, age)

IPEDS variables in IPEDS only

Tested in this study
Enrollment percentage by race-ethnicity and gender
Enrollment percentage of first-time first-year students
Enrollment percentage of full-time students
Enrollment percentage of degrees awarded
Percentage of degrees awarded by race-ethnicity and gender

Other
FTE size categories
Expenditure ratios per FTE
Program offerings by CIP
Accreditation
Fees, tuition, and charges
Categorical expenditures
Number of full-time faculty
Number faculty on tenure/non-tenure
Selected student services
Rural/urban mix
Revenue and expenditure rates
Research expenditures
Funding control
Number of campuses
Mix of high/low cost programs
Credit FTE students/FTE staff
Headcount students/FTE staff
Part-time faculty & Part-time staff/FTE staff

[9] Appendix table 1A provides the same list with some IPEDS variable names.

Some of these variables, such as sector, level of offerings, funding control, accreditation, and admissions requirements are useful for classifying colleges of all levels, but do not differentiate among community colleges very well, so they were not tested. However, 'sector' was used in order to compare the public two-year colleges as a group with the four-year public and private colleges.[10] Other variables listed here are potentially relevant to community colleges, but were not tested as part of this study for various reasons. Some variables (fees/tuition, student services, number of campuses) did not have enough variation or relevance that would distinguish different types of community colleges, while others (program offerings by CIP) had too many categories. Some variables (number of programs, categorical expenditures, research expenditures) had missing data in IPEDS, and for others (FTE size, rural/urban, other expenditure and faculty variables), the data was not accessible to this researcher at the time of the study due to difficulties identifying the values or extracting them from the dataset.[11]

The variables listed in table 1 as "tested in this study" were both accessible and had enough variation to have potential for differentiating community colleges, especially those in California.

The search for peer group variables for the community colleges

IPEDS variables in BPS

In order to identify peer groups that were meaningful to California community colleges, the main criteria for peer group variables was that they would create national peer groups with different levels of racial-ethnic diversity. Racial-ethnic diversity is very salient to the California community colleges because students of color are now a majority of the community college student population in the state. Consequently, in looking for peer groups on a national level, many California colleges look for colleges with as much racial-ethnic diversity as their college. Therefore, the institutional peer group variables in BPS and IPEDS were tested using race-ethnicity as the BPS student-level variable that would be differentiated by the peer groups.

First, the IPEDS variables found in BPS were tested as peer group variables. If these could distinguish among community colleges, then there would be no need to use the separate IPEDS files. BPS contained the variables of sector (control and level), enrollment size, region, state, calendar system, and admissions requirements. Since there was little variation in the calendar system in California or the admissions requirements among public two-year colleges in IPEDS, these variables were not tested.

---

[10]The terms "public two-year colleges," the IPEDS label for these colleges, and "community colleges," are used interchangeably in this report.

[11]For details on the accessibility of IPEDS variables, see the hardware/software and time requirement sections of Chapter IV.

Tables 2 through 5 show the race-ethnicity distribution among the students from colleges grouped by sector, enrollment size, region, and state. Examining the three major sectors—public four-year, private four-year, and public two-year—table 2 shows that racial-ethnic diversity is somewhat higher among public two-year colleges, with white students at 77 percent of the enrollment vs. 82 to 85 percent in the four-year sectors. However, this means that only 23 percent of community college students were people of color, which is far below their representation of over 50 percent of California community college students.

Table 2. Percentage distribution of race-ethnicity of 1989-90 beginning postsecondary students, by selected institutional sectors

	Institutional sectors		
	Public 4-year	Private 4-year	Public 2-year
White	82	85**	77**
African American	8	6	8
Latino	4**	5	10**
Asian/Pacific Islander	5	4	4
Native American	<1	1	1
	100%	100%	100%

NOTE: ** significant difference at p <.01 between values with '**' in same row
SOURCE: U.S. Dept. of Education, OERI, NCES, Beginning Postsecondary Students Longitudinal Study (BPS: 90/94), Restricted-use File

It was hoped that more diversity would be found by focusing on different types of community colleges. Table 3 shows five levels of enrollment size among community colleges, and here the diversity is somewhat higher in the largest colleges. However, white students are still the majority at 63 percent.

Table 3. Percentage distribution of race-ethnicity of 1989-90 beginning postsecondary students in public 2-year colleges, by enrollment size of college

	Enrollment size				
	Under 2,500	2,500-4,999	5,000-9,999	10,000-19,999	20,000 plus
White	86**	84	74	69**	63
African American	9	10	9	5	10
Latino	5	5	12	16	19
Asian/Pacific Islander	0	2	4	9	8
Native American	1	0	1	1	0
	100%	100%	100%	100%	100%

NOTE: ** significant difference at p <.01 between values with '**' in same row
SOURCE: U.S. Dept. of Education, OERI, NCES, Beginning Postsecondary Students Longitudinal Study (BPS: 90/94), Restricted-use File

At this point, just to check that the racial diversity in California was reflected in the NPSAS/BPS sample, diversity was examined in the Far West region and in California and compared to selected regions and states that were the largest and most likely to have diverse students. Tables 4 and 5 show that the student samples in the Far West region and California do indeed reflect the actual diversity of community colleges students in these areas, with 52 and 47 percent white students respectively. However, these tables also show that none of the other likely regions and states match that diversity.

Table 4. Percentage distribution of race-ethnicity of 1989-90 beginning postsecondary students in public 2-year colleges, by selected regions

	Middle East	Great Lakes	South East	Far West
White	83**	97	85	52**
African American	11	7	9	8
Latino	5**	3	4	25**
Asian/Pacific Islander	2	2	1	12
Native American	0	1	<1	2
	100%	100%	100%	100%

NOTE: ** significant difference at p <.01 between values with '**' in same row
SOURCE: U.S. Dept. of Education, OERI, NCES, Beginning Postsecondary Students Longitudinal Study (BPS: 90/94), Restricted-use File

Table 5. Percentage distribution of race-ethnicity of 1989-90 beginning postsecondary students in public 2-year colleges, by selected states

	New York	Texas	Florida	California
White	81**	66	85	47**
African American	10	11	6	9
Latino	8	17	8	29
Asian/Pacific Islander	1	6	1	13
Native American	0	0	0	2
	100%	100%	100%	100%

NOTE: ** significant difference at p <.01 between values with '**' in same row
SOURCE: U.S. Dept. of Education, OERI, NCES, Beginning Postsecondary Students Longitudinal Study (BPS: 90/94), Restricted-use File

This could lead to the conclusion that states are their own best comparison group for the IR variables in BPS, especially in the case of California. However, there are several problems with that solution. Most importantly, the sample sizes in BPS are too small within the states and regions to use those characteristics for peer groups. Appendix tables B1 and B2 provide the standard errors for tables 2 through 5, and they show that as the sample gets smaller, standard errors increase from

2.29 to 4.86 to 9.42 for the percentage of white students, making differences between the peer groups meaningless. Consequently, if colleges wanted to use their own state as a comparison, they would be better off getting the actual data from their state rather than just a sample.[12] In addition, the purpose of using BPS and IPEDS is to find peer groups on a national level that are different from the state comparisons. Therefore, in order to provide these national comparisons, the search for peer group variables was continued in the IPEDS file.

IPEDS variables in IPEDS only

In order to use the IPEDS variables that were only in IPEDS to create peer group variables for the students in BPS, these variables needed to be selected and merged onto the BPS file. The summary section on working with restricted-use data sets in Chapter IV outlines how this was done. The variables that were selected as potential community college peer group variables were enrollment size (to check its consistency with the IPEDS data in BPS), percentage of full-time students, percentage of first-time first-year students, and racial-ethnic diversity (percentage of students of color in enrollments and among first-year students). These variables were categorized into two or three major groups. Of 1,221 public two-year colleges in IPEDS, 218 were in the BPS sample, and 889 students were sampled in these colleges. Table 6 shows the percentage distributions of the peer group variables for the 1,221 public two-year colleges in IPEDS and for the 889 weighted students in BPS.[13]

These variables were chosen because they are factors that reflect student characteristics or affect student outcomes. Enrollment size indicates the number of students on campus regardless of FTE, and it reflects the academic and physical capacity of the college as well as the size of the campus community from a student perspective. Percentage of full-time students is a complement to enrollment size, in that it measures the type of students and intensity of their use of the campus. More full-time students indicate that more students are pursuing transfer, degree, or certificate

---

[12]In California, the Chancellor's Office of the California Community Colleges maintains a statewide database of major IR student characteristics and outcomes from mandatory data provided by all the colleges. The Chancellor's Office provides summary reports based on these data.

[13]The high percentage of missing among these IPEDS variables illustrates one problem with some variables in this data set. When these IPEDS variables were merged onto the file of BPS schools (a subset of the IPEDS schools), similar percentages of schools were still missing these values. However, as shown in table 6, when the BPS school file with new IPEDS variables was merged back onto the BPS weighted student file, the highest percent missing was 9 percent, so the effect of missing values on the estimates was within reasonable limits. In addition to missing values in some IPEDS variables, out of the 218 public two-year schools that were in the BPS sample, only 208 were found in the 1990 IPEDS data set, so these missing schools contributed some missing values. In addition, only 178 of the 208 schools were actually classified as public two-year colleges. NCES staff and IPEDS consultants stated that these mismatches between IPEDS and BPS were to be expected for three reasons: 1) the 1987-88 IPEDS file had been used for the 1990 NPSAS/BPS sampling, 2) colleges change categories over time, and 3) about 5% of the schools in IPEDS are misclassified by level/control (E-mail correspondance with A. Malizio, D. Carroll, and L. Berkner, Oct 31-Nov 5, 1996.) .

Table 6. Percentage distribution of selected peer group variables in 1990 IPEDS population of public two-year colleges and BPS sample of public two-year students

	IPEDS population of public 2-yr colleges	BPS wtd. normalized* sample of public 2-yr students
Total number	*1,221*	*889*
	Percentages	
Enrollment size in October 1990		
Under 2,500	51	17
2,500-9,999	34	43
10,000 or more	10	32
Missing	*4*	*9*
Total	100%	100%
Percentage of full-time students out of total enrollment		
Low: Under 35% full-time students	32	51
High: 35% or more full-time students	47	40
Missing	*21*	*9*
Total	100%	100%
Percentage of first-time first-year students out of total enrollment		
Low: Under 35% freshmen	65	
High: 35% or more freshman	13	not merged
Missing	*22*	into BPS
Total	100%	
Racial-ethnic diversity of enrollment		
Low: Under 20% students of color	50	60
Medium: 20-34% students of color	15	19
High: 35% or more students of color	14	13
Missing	*21*	*9*
Total	100%	100%
Racial-ethnic diversity of first-time first-year students		
Low: Under 20% students of color	45	
Medium: 20-34% students of color	18	not merged
High: 35% or more students of color	15	into BPS
Missing	*22*	
Total	100%	

NOTE: *In normalized samples, the weights have been reduced proportionately so that the cases sum to the actual sample size rather than to the estimated population

SOURCE: U.S. Dept. of Education, OERI, NCES, Integrated Postsecondary Education Data System (IPEDS) 1990 and Beginning Postsecondary Students Longitudinal Study (BPS: 90/94), Restricted-use File.

programs rather than taking just a few courses. The percentage of first-time first-year students (freshmen) can also reflect the type of students and college programs. More freshmen may mean fewer continuing students, which may indicate more students with short term goals. Alternatively, the college may be transferring students soon after the first term. The percentage of racial-ethnic diversity, both in enrollments and among first-time first-year students measures the cultural variation in the college population. Diversity in enrollments and among first-time students was so similar that the more inclusive diversity in enrollments was chosen.

Other variables were considered. The percentage of women did not have enough variation to create categories. Other variables examined were number of AA degrees, percentage of AA degrees out of enrollments, and percentage of women and of racial-ethnic groups among the AA degrees.[14] However, besides the little variation in these variables and the small sample sizes in the gender and race-ethnicity of degrees, these variables were rejected because they were outcome measures, rather than descriptors of the general institutional inputs and characteristics that might lead to student outcomes.

Cross tabulations and chi square analyses were conducted on the selected variables in table 6 for all the public two-year colleges in IPEDS. The goal of these analyses was to identify variables that were related to each other but described different aspects of the colleges. The results, shown in appendix table B4, showed that enrollment size in three categories, percentage full-time in two categories, and percentage first-time in two categories were all highly associated with each other at the $p<.00001$ level. In addition, enrollment size and percentage full-time were associated with levels of diversity (in three and two categories) at the $p<.001$ level, and percentage first-time was not associated with diversity. Since diversity in enrollments, enrollment size, and percentage full-time were mostly closely related, they were chosen as the major peer group variables to be tested with the BPS student characteristics.

Again, due to the salience of racial-ethnic diversity for California community colleges, the BPS student-level variable of racial-ethnic diversity was used as the criteria for how well these institutional variables differentiated among community colleges with different levels of diversity. Not surprisingly, using the student racial-ethnic data, levels of diversity differentiated the most among colleges in the area of racial-ethnic diversity.

---

[14]These tables can be found in appendix table B3.

Table 7 shows the race-ethnicity distribution for schools at three different levels of diversity. Even though the peer group with the highest level of diversity is only 35 percent or more students of color, only 31 percent of students at these colleges are white, while 69 percent are students of color, so this category does identify diverse colleges. Due to low sample sizes, it was apparent that two categories would be better than three categories for peer variables. Therefore, table 8 shows the race-ethnicity distribution for schools at only two different levels of diversity. With the highest peer group category of diversity at 20 percent or more, these colleges still had 49 percent students of color, which is close to the California average of 51 percent.

Table 7. Percentage distribution of race-ethnicity of 1989-90 beginning postsecondary students in public 2-year colleges, by percentage of students of color in student body

	Level of diversity: percentage of students of color		
	Low	Medium	High
	≤20%	20-35%	≥35%
White	89	64	31
African American	5	12	18
Latino	4	17	36
Asian/Pacific Islander	1	5	15
Native American	≤1	2	1
	100%	100%	100%

SOURCE: U.S. Dept. of Education, OERI, NCES, Beginning Postsecondary Students Longitudinal Study (BPS: 90/94), Restricted-use File

Table 8. Percentage distribution of race-ethnicity of 1989-90 beginning postsecondary students in public 2-year colleges, by percentage of students of color in student body

	Level of diversity: percentage of students of color	
	Low	High
	≤20%	≥20%
White	89	51
African American	5	15
Latino	4	25
Asian/Pacific Islander	1	9
Native American	≤1	1
	100%	100%

SOURCE: U.S. Dept. of Education, OERI, NCES, Beginning Postsecondary Students Longitudinal Study (BPS: 90/94), Restricted-use File

Tables 9 and 10 show the race-ethnicity distribution for schools in the enrollment size and percentage full-time categories. While the race-ethnicity differences are not as pronounced with these variables, they do reflect some variation in diversity. Since these variables were associated with diversity, they take diversity into account while expressing college characteristics that might be more salient than diversity to some colleges. Therefore, all three of these peer group variables were used in estimating student characteristics and outcomes.[15]

Table 9. Percentage distribution of race-ethnicity of 1989-90 beginning postsecondary students in public 2-year colleges, by enrollment size of college

	Enrollment size		
	Under 2,500	2,500-9,999	10,000-plus
White	86	78	68
African American	9	9	6
Latino	5	10	16
Asian/Pacific Islander	0	3	9
	100%	100%	100%

SOURCE: U.S. Dept. of Education, OERI, NCES, Beginning Postsecondary Students Longitudinal Study (BPS: 90/94), Restricted-use File

Table 10. Percentage distribution of race-ethnicity of 1989-90 beginning postsecondary students in public 2-year colleges, by percentage of full-time students in student body

	Percentage of full-time students	
	Low ≤35%	High ≥35%
White	68	86
African American	10	7
Latino	15	6
Asian/Pacific Islander	6	1
Native American	1	1
	100%	100%

SOURCE: U.S. Dept. of Education, OERI, NCES, Beginning Postsecondary Students Longitudinal Study (BPS: 90/94), Restricted-use File

[15]Standard errors for these tables are included in appendix tables C12a and C12b, and significant differences are discussed in Chapter III.

Conclusion

While regional and state groups may be the best comparison groups for many community colleges, sample sizes in BPS necessitate finding peer group variables that can group all community colleges nationally into only two or three groups. Not surprisingly, the percentage of students of color differentiated most among colleges in terms of the race-ethnicity of the students. Therefore, this diversity variable in two categories (above and below 20 percent students of color), was chosen as the best peer group variable for California and other community colleges for whom diversity is important.

However, several other variables emerged as correlated to diversity levels that might be important to other community colleges. These were the percentage of full-time students in two categories (above and below 35 percent) and enrollment size in three categories (under 2,500, 2,500-9,999, and 10,000 or over). Due to sample sizes, it was not possible to combine these three variables to create more refined peer groups. Therefore, tables of major BPS student characteristics and outcomes are shown and discussed separately for peer groups based on categories of diversity, percentage full-time and enrollment size. Community colleges are encouraged to use the peer group category that is most salient for them.

Chapter III

Usefulness
Finding IR student characteristics and outcomes variables in BPS

III. Usefulness
Finding IR student characteristics and outcomes variables in BPS

Common student variables used by community colleges

One of the major components of institutional research reports in community colleges, and in most other colleges, is the reporting of the demographics, enrollment patterns, and outcomes of the students in that college.[16] Having comparable data on these student characteristics and outcomes for peer group colleges increases the meaningfulness of the college data.[17] The purpose of this report is to provide those peer group comparisons for community colleges on the most commonly-reported student characteristics and outcomes that are also found in the BPS dataset.

The student characteristics reported by most colleges are basic demographics (gender, race-ethnicity, age, disability, city or high school of origin, current residence) and enrollment status (new, transfer, or continuing student, full-time/part-time status, day/evening, current educational level, major, educational goal).[18] Some colleges also collect and report information on student and family income, student financial dependency status, parents' educational level, marital status, religion, hours of paid work, current occupation, and future aspirations.[19]

The major student outcomes reported in community colleges are the course retention, success, and withdrawal rates, persistence rates, numbers and rates of transfers, degrees, and certificates, and time to degree or transfer.[20]

---

[16]Larry G. Jones, "A Brief History of the Fact Book," *New Directions for Institutional Research*, No 91, Fall 1996 and Jean J. Endo, "Developing the Contents of Institutional Fact Books," *New Directions for Institutional Research*, No 91, Fall 1996.

[17]Marks, op. cit.

[18]Endo, op. cit.

[19]*1993-94 Student Survey*, Los Angeles Community Colleges, Los Angeles, CA; *Chabot College Campus Climate Survey, Fall 1994*, Chabot College, Hayward, CA.

[20]*The Effectiveness of California Community Colleges on Selected Performance Measures*, Chancellor's Office, California Community Colleges, Sacramento, CA, October 1996.

IR student characteristics and outcomes variables in BPS

The BPS dataset contains most of the commonly-reported student characteristics and outcomes variables. With its base in NPSAS, it is especially strong on the details of financial status, financial aid, and family data. However, because BPS is a longitudinal study, it also contains many variables on persistence and transfer. Table 11 lists the student variables available in BPS that institutional researchers would be most interested in.[21] The major demographics and enrollment status variables are on this list, with the exception of day or evening attendance and the city and high school of origin. The strongest student outcomes variables are the persistence variables, but degree and transfer attainment are also available. While grades are included, they do not include withdrawals, so they are not comparable to community college success/withdrawal rates. However, most of the basic institutional research student characteristics and outcomes can be estimated with the BPS dataset.

Table 11. Types of student characteristics and outcomes variables available in BPS

Student characteristics
Gender
Race-ethnicity
Citizenship
Religion
Age
Location of residence in relation to college
Type of HS/Time between high school and college
Educational goal
Major/college programs
Full/part-time
Type and amount of college attendance
Experience in basic skills
Family SES, educational and occupational background
Level of work, income, and financial support
Family obligations
Overall risk factors

Student outcomes
Grades
First year retention/persistence
First through four year transfer and degree attainment

[21] The BPS variable names are listed in appendix table A3.

Student characteristics and outcomes variables, by sector and peer groups

Tables 12a through 16b present estimates of the major student characteristics and outcomes found in BPS for public two-year colleges. First, in order to provide a comparison for community colleges, these estimates are shown for the three sectors of interest—public four-year colleges, private four-year colleges, and public two-year colleges. Then, these tables show the same estimates for public two-year colleges only, divided into two levels of diversity (percentage of students of color), three categories of enrollment size, and two levels of percentage of full-time students. Significant differences between categories are noted. Standard errors are shown in corresponding appendix tables C12a to C16b. Standard errors for sector and enrollment were calculated using DAS. Standard errors for diversity and full-time status were estimated using design effects. For more details on these estimates and tests, see the sections on estimating standard errors and testing for differences in Chapter IV. Examples of calculating standard errors using design effects and testing for differences using standard errors are provided in appendix table C1.

Table 12a. Percentage distribution of gender, race-ethnicity, citizenship, age, and religion of beginning postsecondary students, by sector and by percentage of students of color within public two-year colleges, 1989-90

	Sector			Public two-yr colleges Percentage of students of color	
	Public 4-yr	Private 4-yr	Public 2-yr	Under 20%	20% or more
Gender					
Male	46	49	48	48	50
Female	54	51	52	52	50
Race-ethnicity					
White, non-Latino	82	85**	77**	89**	51**
African-Am, non-Latino	8	6	8	5**	15**
Latino	4**	5	10**	4**	25**
Asian/Pacific Islander	5	4	4	1**	9**
Native American	<1	1	1	<1	1
U.S. Citizenship					
U.S. Citizen	98	98	97	99	93
Not a U.S. Citizen	2	2	3	1	7
Age					
19 yrs or less	91**	92**	63**	64	60
20-24 yrs	5	4	16	14	20
25-29 yrs	1	1	7	7	8
30-39 yrs	2	2	8	8	7
40 yrs or more	1	1	6	7	6
Religion					
Protestant	39	31	34	37**	27**
Catholic	34	38	32	31	39
Jewish	1	4	<1	1	0
Other religion	17	20	21	20	22
None	9	8	12	12	14

NOTES: Columns within each student variable add to 100 percent (or close to 100 percent due to rounding).
Significant difference at p< .01 between values with '' in same row and peer group variable
SOURCE: U.S. Dept. of Education, OERI, NCES, Beginning Postsecondary Students Longitudinal Study (BPS: 90/94), Restricted-use File

Table 12b. Percentage distribution of gender, race-ethnicity, citizenship, age, and religion of beginning postsecondary students in public two-year colleges, by enrollment size and percentage of full-time students: 1989-90

	Enrollment size			Percentage of full-time students	
	Under 2,500	2,500 to 9,999	10,000 or more	Under 35%	35% or more
Gender					
Male	49	46	50	46	52
Female	51	54	50	54	48
Race-ethnicity					
White, non-Latino	86**	78	68**	68**	86**
African-Am, non-Latino	9	9	6	10	7
Latino	5**	10	16**	10	7
Asian/Pacific Islander	0	3**	9**	6	1
Native American	1	1	1	1	1
U.S. Citizenship					
U.S. Citizen	99	98	95	95	100
Not a U.S. Citizen	2	2	5	5	<1
Age					
19 yrs or less	67	60	68	60	66
20-24 yrs	12	15	19	19**	12**
25-29 yrs	8	8	4	7	7
30-39 yrs	6	9	5	8	8
40 yrs or more	7	8	3	6	8
Religion					
Protestant	47	34	27	30	39
Catholic	20	33	38	35	31
Jewish	0	1	<1	0	1
Other religion	26	19	21	20	22
None	7	14	14	15	9

NOTES: Columns within each student variable add to 100 percent (or close to 100 percent due to rounding).
Significant difference at p< .01 between values with '' in same row and peer group variable
SOURCE: U.S. Dept. of Education, OERI, NCES, Beginning Postsecondary Students Longitudinal Study (BPS: 90/94), Restricted-use File

Table 13a. Percentage distribution of SES, parent's education, high school diploma and years since high school, marital status and number of children of beginning postsecondary students, by sector and by percentage of students of color within public two-year colleges, 1989-90

	Sector			Public two-yr colleges Percentage of students of color	
	Public 4-yr	Private 4-yr	Public 2-yr	Under 20%	20% or more
Socio-economic status (SES)					
Lowest quartile	6**	5**	19**	18	21
Low middle quartile	11	9	21	17	25
High middle quartile	29	23	30	32	30
Highest quartile	55**	63**	31**	33	25
Parent's highest education					
High school or less	31	26	51	52	50
Trade school	5	5	3	3	2
Some college	21	17	18	17	19
Bachelor's degree	24	24	17	18	18
Graduate/prof. degree	19	28	10	9	11
Type of high school diploma					
Regular diploma	99	98	92	93	92
Other or none	1	2	8	7	8
Number of years since high school graduation					
Same year	89	90	61	60	59
1-2 years	6	4	13	13	16
3-9 years	3	3	11	11	11
10 years or more	2	3	15	16	15
Marital status					
Not married	96	96	80	79	80
Married	4	4	19	20	18
Separated	<1	<1	1	1	2
Number of children					
None	97	97	81	80	83
One	2	1	8	8	8
Two or more	2	2	11	12	9
Single parent					
Yes	2	2	6	6	7
No	98	98	94	94	93

NOTES: Columns within each student variable add to 100 percent (or close to 100 percent due to rounding).
Significant difference at p< .01 between values with '' in same row and peer group variable
SOURCE: U.S. Dept. of Education, OERI, NCES, Beginning Postsecondary Students Longitudinal Study (BPS: 90/94), Restricted-use File

Table 13b. Percentage distribution of SES, parent's education, high school diploma and years since high school, marital status and number of children of beginning postsecondary students in public two-year colleges, by enrollment size and percentage of full-time students: 1989-90

	Enrollment size			Percentage of full-time students	
	Under 2,500	2,500 to 9,999	10,000 or more	Under 35%	35% or more
Socio-economic status (SES)					
Lowest quartile	19	20	15	20	17
Low middle quartile	22	21	18	22	18
High middle quartile	27	31	32	30	32
Highest quartile	32	28	35	29	33
Parent's highest education					
High school or less	59	52	44	50	53
Trade school	6	3	2	3	3
Some college	16	17	22	18	19
Bachelor's degree	15	17	18	19	16
Graduate/prof. degree	4	11	14	10	9
Type of high school diploma					
Regular diploma	94	92	94	92	92
Other or none	6	8	6	8	8
Number of years since high school graduation					
Same year	66	58	65	58	62
1-2 years	7	14	17	15	13
3-9 years	14	11	9	12	10
10 years or more	13	18	10	15	16
Marital status					
Not married	77	76	88	79	80
Married	20	23	10	19	19
Separated	3	1	1	2	1
Number of children					
None	79	79	87	80	83
One	9	9	5	8	7
Two or more	12	12	8	12	10
Single parent					
Yes	8	6	4	6	6
No	92	94	96	94	94

NOTES: Columns within each student variable add to 100 percent (or close to 100 percent due to rounding).
Significant difference at p< .01 between values with '' in same row and peer group variable
SOURCE: U.S. Dept. of Education, OERI, NCES, Beginning Postsecondary Students Longitudinal Study (BPS: 90/94), Restricted-use File

Table 14a. Percentage distribution of distance from home, residence, financial dependency, college attendance, and major of beginning postsecondary students, by sector and by percentage of students of color within public two-year colleges, 1989-90

	Sector			Public two-yr colleges Percentage of students of color	
	Public 4-yr	Private 4-yr	Public 2-yr	Under 20%	20% or more
Distance of college from home					
5 miles or less	10	8	26	23	32
6-10 miles	10	6	23	21	30
11-50 miles	31	24	46	49**	35**
51-100 miles	15	13	2	3	2
101-500 miles	29	32	3	4	1
Over 500 miles	5	17	<1	<1	<1
Type of residence					
Campus housing	56	76	3	4	2
Off-campus	15	8	40	39	40
With parents	29	16	57	57	58
Financial dependency status					
Dependent	91	92	67	66	69
Independent	9	8	33	35	32
College attendance status					
Full-time	89	94	48	53**	38**
Half-time	8	4	26	25	31
Less than half-time	3	3	26	23	31
Major in 1989-90					
Humanities	13	17	17	16	19
Social and behav. sci.	14	20	4	3	7
Life sciences	6	7	4	4	4
Physical sciences	2	2	<1	<1	0
Mathematics	2	1	<1	1	1
Computer and info sci	<1	<1	<1	<1	1
Engineering	11	8	11	10	12
Education	11	8	7	7	6
Business & mgmnt.	21	23	27	27	26
Health	8	6	11	12	8
Vocational/tech.	3	1	9	9	9
Other technical	9	6	8	10	7

NOTES: Columns within each student variable add to 100 percent (or close to 100 percent due to rounding).
Significant difference at p< .01 between values with '' in same row and peer group variable
SOURCE: U.S. Dept. of Education, OERI, NCES, Beginning Postsecondary Students Longitudinal Study (BPS: 90/94), Restricted-use File

Table 14b. Percentage distribution of distance from home, residence, financial dependency, college attendance, and major of beginning postsecondary students in public two-year colleges, by enrollment size and percentage of full-time students: 1989-90

	Enrollment size			Percentage of full-time students	
	Under 2,500	2,500 to 9,999	10,000 or more	Under 35%	35% or more
Distance of college from home					
5 miles or less	17	28	28	27	26
6-10 miles	20	22	28	30**	17**
11-50 miles	50	45	42	41	49
51-100 miles	5	2	1	1	3
101-500 miles	8	2	2	1	6
Over 500 miles	0	1	0	<1	0
Type of residence					
Campus housing	11	2	<1	1	6
Off-campus	41	44	29	41	38
With parents	47	54	70	58	57
Financial dependency status					
Dependent	66	62	76	66	67
Independent	34	38	24	34	33
College attendance status					
Full-time	69	43	43	39**	59**
Half-time	19	25	36	32**	20**
Less than half-time	12	33	22	29	21
Major in 1989-90					
Humanities	14	17	20	20	14
Social and behav. sci.	4	4	7	5	4
Life sciences	2	5	4	4	4
Physical sciences	1	0	0	0	<1
Mathematics	1	0	1	<1	1
Computer and info sci	0	1	1	1	0
Engineering	15	13	7	9	14
Education	4	8	6	7	7
Business & mgmnt.	31	23	29	25	29
Health	11	11	11	12	8
Vocational/tech.	8	9	9	9	9
Other technical	9	9	5	7	11

NOTES: Columns within each student variable add to 100 percent (or close to 100 percent due to rounding).
Significant difference at p< .01 between values with '' in same row and peer group variable
SOURCE: U.S. Dept. of Education, OERI, NCES, Beginning Postsecondary Students Longitudinal Study (BPS: 90/94), Restricted-use File

Table 15a. Percentage distribution of work hours, current occupation, and future occupation goal of beginning postsecondary students, by sector and by percentage of students of color within public two-year colleges, 1989-90

	Sector			Public two-yr colleges Percentage of students of color	
	Public 4-yr	Private 4-yr	Public 2-yr	Under 20%	20% or more
Number of hours worked for pay in 1989-90					
None	22	22	17	17	16
1-19 hours	21	29	13	12	13
20-34 hours	33	24	31	30	32
35 or more hours	25	24	39	41	38
Current occupation*					
Professional	5	8	6	6	5
Executive	4	3	7	7	7
Marketing	27	22	21	18	26
Administrative support	19	19	20	19	20
Technical	2	2	2	1	3
Service	26	27	24	27	22
Blue collar	19	18	20	22	17
Occupation expected in future					
Marketing	6	6	5	5	3
Administrative support	4	5	9	9	9
Service	3	3	7	8	7
Executive	20	23	21	21	24
Postsecondary teacher	1	1	<1	<1	0
Other education	16	14	9	9	7
Engineer/Architect	11	8	8	7	9
Scientist	4	5	2	2	2
Computer science	2	1	3	3	3
Social/recreation	2	4	2	3	2
Doctor/dentist	3	3	1	1	1
Other medical	9	6	12	12	12
Technical	3	2	5	5	4
Lawyer	3	5	2	1	4
Blue collar	4	4	7	8	6
Writer/artist	11	11	7	6	9

NOTES: Columns within each student variable add to 100 percent (or close to 100 percent due to rounding).
Significant difference at p< .01 between values with '' in same row and peer group variable
*Variable not available in DAS so standard errors and difference tests could not be calculated
SOURCE: U.S. Dept. of Education, OERI, NCES, Beginning Postsecondary Students Longitudinal Study (BPS: 90/94), Restricted-use File

Table 15b. Percentage distribution of work hours, current occupation, and future occupation goal of beginning postsecondary students in public two-year colleges, by enrollment size and percentage of full-time students: 1989-90

	Enrollment size			Percentage of full-time students	
	Under 2,500	2,500 to 9,999	10,000 or more	Under 35%	35% or more
Number of hours worked for pay in 1989-90					
None	15	17	17	16	18
1-19 hours	16	12	13	12	14
20-34 hours	36	29	33	34	28
35 or more hours	33	43	37	39	41
Current occupation*					
Professional	4	6	5	7	4
Executive	5	9	5	5	9
Marketing	20	17	28	22	18
Administrative support	19	22	19	19	21
Technical	1	2	2	2	2
Service	27	23	26	27	23
Blue collar	24	20	15	17	24
Occupation expected in future					
Marketing	3	4	5	6	3
Administrative support	11	9	8	7	11
Service	3	10	6	8	7
Executive	23	19	23	23	19
Postsecondary teacher	0	1	0	0	1
Other education	7	10	9	7	10
Engineer/Architect	9	6	10	10	5
Scientist	3	2	3	3	1
Computer science	5	4	1	2	4
Social/recreation	3	2	3	2	2
Doctor/dentist	0	2	1	<1	2
Other medical	11	12	11	15	9
Technical	8	5	3	3	7
Lawyer	3	1	4	2	2
Blue collar	10	7	3	5	11
Writer/artist	4	7	10	7	7

NOTES: Columns within each student variable add to 100 percent (or close to 100 percent due to rounding).
Significant difference at p< .01 between values with '' in same row and peer group variable
*Variable not available in DAS so standard errors and difference tests could not be calculated
SOURCE: U.S. Dept. of Education, OERI, NCES, Beginning Postsecondary Students Longitudinal Study (BPS: 90/94), Restricted-use File

Table 16a. Percentage distribution of degree goal, educational aspirations, persistence and attainment, and degree progression of beginning postsecondary students, by sector and by percentage of students of color within public two-year colleges, 1989-90

	Sector			Public two-yr colleges Percentage of students of color	
	Public 4-yr	Private 4-yr	Public 2-yr	Under 20%	20% or more
Degree currently working toward					
None	2	1	7	8	8
Vocational certificate	1	1	13	11	15
Associate's degree	6	6	54	59	48
Bachelor's degree	91	91	26	23	30
Educational aspirations					
Trade school	1	1	7	7	6
Associate's degree	3	2	21	24	19
Bachelor's degree	37	29	41	43	41
Advanced degree	60	68	31	27	33
Persistence and attainment in first year					
Attained Certificate	<1	<1	2	2	1
Persisted to next year	72	73	45	45	46
Transferred during year	4	5	4	2	3
Subsequent transfer	13	15	17	17	19
Stopout in 89/90	4	3	11	10	15
Left without returning	6	4	21	24	17
Degree progression in 4 years					
Attained BA	46	65	3	3	3
Attained AA, then BA	1	2	4	4	4
Attained AA, enrld BA	1	<1	2	2	1
Attained AA, not enrld BA	4	3	15	16	15
Attained certificate, then degree or enrolled	1	1	1	1	0
Attained cert, not enrolled	3	2	13	12	12
No degree, enrolled BA	16	7	7	5	8
No degree, enrolled AA	2	1	6	5	9
No degree, enrolled cert	<1	<1	1	1	1
No degree, not enrolled	27	20	49	51	46
Degree progression in 4 years: Summary					
Attained BA	47	67	7	7	7
Attained AA	5	4	17	18	16
Attained certificate,	4	3	14	13	12
No degree, enrolled	19	9	14	11	18
No degree, not enrolled	27	20	49	51	46

NOTES: Columns within each student variable add to 100 percent (or close to 100 percent due to rounding).
Significant difference at p< .01 between values with '' in same row and peer group variable
SOURCE: U.S. Dept. of Education, OERI, NCES, Beginning Postsecondary Students Longitudinal Study (BPS: 90/94), Restricted-use File

Table 16b. Percentage distribution of degree goal, educational aspirations, persistence and attainment, and degree progression of beginning postsecondary students in public two-year colleges, by enrollment size and percentage of full-time students: 1989-90

	Enrollment size			Percentage of full-time students	
	Under 2,500	2,500 to 9,999	10,000 or more	Under 35%	35% or more
Degree currently working toward					
None	6	9	6	8	7
Vocational certificate	12	15	9	12	12
Associate's degree	60	56	51	56	54
Bachelor's degree	22	21	35	24	27
Educational aspirations					
Trade school	10	7	3	5	9
Associate's degree	23	24	15	21	23
Bachelor's degree	42	39	46	44	39
Advanced degree	26	30	35	29	29
Persistence and attainment in first year					
Attained Certificate	2	2	<1	<1	3
Persisted to next year	51	41	50	45	46
Transferred during year	4	4	3	3	2
Subsequent transfer	14	18	17	18	17
Stopout in 89/90	5	12	12	13	10
Left without returning	24	23	17	21	23
Degree progression in 4 years					
Attained BA	5	2	4	4	3
Attained AA, then BA	6	3	4	5	4
Attained AA, enrld BA	3	3	2	1	3
Attained AA, not enrld BA	17	14	15	14	16
Attained certificate, then degree or enrolled	1	1	0	0	2
Attained cert, not enrolled	11	14	10	14	9
No degree, enrolled BA	4	6	10	6	6
No degree, enrolled AA	2	7	8	7	6
No degree, enrolled cert	<1	2	1	1	1
No degree, not enrolled	51	48	46	49	50
Degree progression in 4 years: Summary					
Attained BA	11	5	8	9	7
Attained AA	20	17	17	15	19
Attained certificate,	12	15	10	14	11
No degree, enrolled	7	14	19	14	13
No degree, not enrolled	51	48	46	49	50

NOTES: Columns within each student variable add to 100 percent (or close to 100 percent due to rounding).
Significant difference at p< .01 between values with '' in same row and peer group variable
SOURCE: U.S. Dept. of Education, OERI, NCES, Beginning Postsecondary Students Longitudinal Study (BPS: 90/94), Restricted-use File

Differences by sector

Many of the largest differences in student characteristics and outcomes were between the public two-year sector (community colleges) and the public and private four-year sectors. The major differences in the variables from gender to SES in tables 12a and 13a were tested, and these tests showed that most differences over 8 percentage points were significant. Compared to public or private four-year colleges, community colleges have smaller proportions of white students and much smaller proportions of students age 19 years or less (table 12a). In addition, community college students are much more likely to be from the lowest SES quartile, and much less likely to be from the highest SES quartile (table 13a). Tables 13a through 16a show many other large differences between community colleges and the four-year colleges, and these differences can be tested using the standard errors provided in tables C13a through C16a.

Differences by diversity

Although diversity was a key peer group variable for public two-year colleges in this study, there were very few differences in student characteristics or outcomes between students in campuses with low and high diversity levels. As expected and discussed earlier, higher percentages of African American, Latino, and Asian students were in colleges of higher diversity. All other major differences were tested, and only the following were significant. In more diverse colleges, students were less likely to be Protestant, to live more than 11 miles away from home, and to attend full-time.

Differences by enrollment

Among public two-year colleges, enrollment size produced some differences in student characteristics and outcomes, but not many. Comparing small (under 2,500) and large (10,000 and over) colleges, there were lower proportions of whites and higher proportions of Latinos and Asians in larger colleges. However, there were no significant differences in SES. In other student characteristics and outcomes, some differences appeared between larger colleges and the other two smaller colleges sizes. These differences can be tested using the standard errors provided in tables C12b to C16b.

Differences by percentage of full-time students

Like the diversity categories, dividing the public two-year colleges into those below and above 35 percent full-time students yielded few significant differences between these two groups. All major differences were tested and only the following significant differences were found. Colleges with 35 percent or more full-time students had higher proportions of white students, lower proportions of students 20-24 years, lower proportions of students living 6-10 miles from campus, and, as would be expected, higher proportions of full-time students and lower proportions of half-time students.

Overall findings

While the largest differences among the commonly-used IR student characteristics and outcomes occurred between sectors, there were some differences among community colleges based on the peer group levels of diversity, enrollment size, and percentage full-time. Classification-based groups cannot often attain the uniqueness of peer groups created with more qualitative data. However, the similarities as well as the differences among these different types of community colleges can be reassuring and helpful for colleges who want to use a national comparison group that reflects their own salient characteristics.

Chapter IV

Accessibility
How to obtain and work with NCES Data sets

IV. Accessibility
How to obtain and work with NCES Data sets

Types of data sets available

IPEDS public-use files

All of the IPEDS files are public-use files, which means that they are easily accessible to the public on request either on CDs or as files on the World Wide Web.[22] IPEDS consists of yearly raw data on the population of all postsecondary (PSE) institutions in the U.S. Since 1990, NCES has provided the IPEDS files each year on free CDs that contain the raw data, data retrieval and analysis software, technical and user manuals, and summary reports. These files are also available on the Web.

Each IPEDS data set consists of five files, some from the latest year (the year on the title of the data set), and some from the previous year, depending on the type of data collected. Three of the files—Institutional Characteristics, Faculty Salaries, and Fall Enrollments—contain data from the latest year, while two of the files—Finance and Completions—consist of data from the previous year. These files are all linked by an institutional ID. In addition, NCES has created a sixth file of all the PSE institutions, called the Analysis file, which contains the most commonly-used variables from each file.[23]

NPSAS and BPS

The NPSAS and BPS data sets are available as either public-use files or restricted-use files. The NPSAS data set consists of a cross-sectional sample of all PSE students in the U.S., including both undergraduates and graduates. These data were collected in 1989-90 and again in 1992-93. The BPS sample of all first-year/first-time undergraduate PSE students in the U.S. is a subsample of the 1989-90 NPSAS. The BPS sample then became a longitudinal data set, and data were collected on these students every year or so for four years. The data from each wave are available either as separate files or as one cumulative file. Both the NPSAS and BPS samples are national samples, which means they do not provide adequate sample sizes for individual state analysis.[24]

---

[22]See Appendix D for ordering and downloading information.

[23]For a discussion of some difficulties in using the Analysis file, see the section on time requirements for public-use files in this chapter.

[24]The NPSAS and BPS files are on separate CDs, but the process of working with them is the same, so they will be discussed together in this report.

NPSAS and BPS public-use files

NPSAS and BPS public-use files come with software that can generate most of the possible frequencies, cross tabulations, and correlation matrices in these data sets. However, access to the raw data is not provided. By not providing the raw data, NCES not only protects the privacy of the respondents, but makes the data analysis process easier for the users.

Like IPEDS, these public-use files are available on request either on CDs or on the Web. The files include the data, which can only be accessed by the enclosed data analysis software, called Data Analysis System (DAS), the user and technical manuals, and summary reports. The advantages of using the public-use files is that DAS is easy to learn and use. Almost immediately, a researcher can produce weighted frequencies, percentages, means, or correlations on the total sample or on subsets such as public two-year colleges, using any variable as a control variable. DAS can be used to create tables of any specifications, print out the tables, and provide the estimated population sizes and standard errors. Since the NPSAS/BPS data sets have a complex sampling design, the standard errors need to be calculated in a way that takes this design into account, and DAS provides these accurate standard errors. The limitations of using the public-use files are that actual sample sizes and totals cannot be seen, variables from other data sets such as IPEDS cannot be added to the data sets, and variables cannot be recoded, combined, or created.[25] For that type of access to the raw data, the restricted-use files are necessary.

NPSAS and BPS restricted-use files

The restricted-use files are provided on CDs and consist of the raw data, an electronic codebook (ECB), SAS and SPSS programs for reading the data, technical and user manuals and summary reports. The advantages of using these files are that researchers can manipulate the raw data in such ways as recoding/combining/creating variables, obtaining all frequencies and statistics, examining the sample sizes, and matching the file to IPEDS data on the sampled institutions.

However, this freedom to use the raw data incurs several costs in time and effort. First, in order to ensure the protection of privacy of the respondents, it is necessary to develop a security plan and obtain a restricted data license from NCES. It takes time to set up and document a security plan and obtain the license. Details of this process are provided later in this chapter under the section on how to obtain restricted-use data. Learning to use the ECB is another time investment. The ECB must be used to choose variables and provide frequencies, a codebook, and SAS or SPSS programs to read the data. In BPS, variable information is found only in the ECB rather than in printed or computer text files, so variables must be reviewed on the computer screen before selecting for inclusion in a custom codebook. Since BPS has hundreds of variables to choose

---

[25]One exception is that consecutive values of discrete or continuous variables can be recoded by grouping into smaller groups and re labeling.

from, this codebook, and the printing costs associated with it, can become rather large. Finally, the SPSS or SAS programs must be modified and run to the researcher's specifications, and the results need to be checked against either the public-use (DAS) results or published sources in order to insure their accuracy.

Preliminary considerations: hardware/software, time, funds

Hardware/software requirements

Public Use Files

In order to use the software provided with the IPEDS and NPSAS/BPS public-use files, a PC with a CD drive and a DOS or Windows operating system is needed. The DAS software is the only mechanism with which to view the NPSAS/BPS variables and create tables. However, since the IPEDS files contain raw data and text codebooks, either a Macintosh or a PC with a CD drive and SPSS-like software can be used to read these files directly from the CD or copy or download them into a text or ASCII file. However, the IPEDS software (a CD-ROM program interface) provides a more complete codebook than the text files, so DOS/Windows is also an advantage for using IPEDS. For the public-use files, these are the minimum requirements:

Hardware:	• NPSAS/BPS:	DOS or Windows operating system; CD drive
	• IPEDS:	Macintosh *or* DOS/Windows with CD drive
Software:	• NPSAS/BPS:	None needed: software (DAS) provided with files
	• IPEDS:	Macintosh: SPSS-like software to read data
		DOS/Windows: software provided with files
Space:	• 500k for minimal installation of DAS	
	• Up to 2 MB for full installation of DAS	
	• Room for DAS result files: 1 to 5k.	
	• Room for IPEDS SPSS program and system files: 10k to 1 MB	
Memory:	• At least 480K memory to run minimum tables on PC with DAS.	

Restricted-Use Files

Since the NPSAS/BPS restricted-use files include the raw data on the CD, any computer with a CD drive that can read ASCII or text files can access the data from the CD. However, in order to know the data file layout and variable names and labels, the Electronic Codebook (ECB) must be used, and this software can only be used on a DOS or Windows operating system. Therefore, all users have to start on a DOS or Windows machine in order to identify the variables and file layout, although the Macintosh users can return to their computers for the analysis step.

On the DOS/Windows computer, the ECB software provides a view of all the variables, their labels, and unweighted frequency/percentage distributions. By selecting variables from this

long list, the user can ask ECB to create a codebook of these variables for reading or printing. ECB will also create an SPSS or SAS program to retrieve those variables from the CD for analysis.

DOS/Windows users can then run these SPSS or SAS programs, which will access the data file on the CD. Macintosh users can copy the codebook and SPSS files onto a floppy disk and transport it to their Mac. There, they can read or print the codebook and edit and run the SPSS programs, which will access the data file on their CD. In order to produce the correct standard errors, both types of computer users will have to obtain additional software or the standard errors will need to be estimated using a spreadsheet.[26] Therefore, for the restricted-use files, these are the minimum requirements:

Hardware:	• PC users:	DOS or Windows operating system; CD drive
	• Macintosh users:	DOS/Windows and Mac computers with CD drives
Software:	• PC users:	ECB provided with files; SPSS or SAS
	• Macintosh users:	SPSS
Space:	• 500k-2 MB to create ECB-generated codebook and SPSS/SAS programs	
	• Room for SPSS/SAS program and system files	
Memory:	• As much as is needed to run SPSS, SAS, or other program	

Time requirements

Public-use files

The Data Analysis System (DAS) that comes with the public-use files of NPSAS, BPS, and other NCES sample data sets is very straightforward to install and learn, and delightfully easy to use. DAS can be installed and learned within an hour, and the first tables can be generated soon after that. After that, the only time constraint is the number of variables and tables the researcher wants to explore and generate.

The IPEDS public-use file also includes software (CD-ROM program interface) that makes the variable viewing, data retrieval, and table generation more automatic, although it was not used in this study because it was only available for DOS/Windows systems. The convenience of reading the IPEDS data directly from the CD into the Macintosh SPSS program was more compelling than using the IPEDS software, especially since this raw data would be merged onto the BPS file. However, this researcher found that it takes more time to use the IPEDS files with a Macintosh, because the variable layouts of all six files must be examined on the screen or in print, and then SPSS programs must be generated from the layout information (although not completely from scratch since the variable formats and variable and value labels can be copied from these files into SPSS programs). Another time expense resulted from the fact that the IPEDS Analysis file was written onto a file of 8,000 plus columns. While the CD-ROM program interface can read this file,

[26]For details, see the section later in this chapter on estimating standard errors.

SPSS for a Macintosh cannot read past 1024 columns. Therefore, programmers had to write a program to chop up the file into 9 records so SPSS could read it, and formulating this new layout took time. This layout is provided in appendix table A2.

Restricted-use files

Restricted-use files take much more time to use. Some of the time is necessary in order to obtain a restricted data license and the data sets. Other time is required to familiarize oneself with the data sets and the ECB software and to insure the accuracy of the results. The rest is the usual time it takes to conduct any original programming and analysis. Time is required to:

1) Create a security plan, get signatures, and obtain security clearance for a restricted data license (allow at least one month)
2) Identify and obtain the correct data sets (if one is using BPS and/or NPSAS and wants to merge the correct IPEDS data with NPSAS or BPS data)
3) Select variables/weights with ECB, edit SPSS/SAS programs, conduct exploratory runs
4) Check programs/results by matching to published reports, DAS tables.
5) Merge IPEDS variables with restricted-use data set
6) Examine results, create new variables, explore different results
7) Estimate standard errors for complex sampling design and test for differences (For various options, see section in this chapter on estimating standard errors)

Funds/supplies/computer support requirements

Although the data sets on the CD's and the Web site are free, other costs are generated by using these data sets, primarily in the area of printing and paper supplies. Paper (either computer or laser) and other printer supplies are needed in order to print out the technical and user manuals, layout records, and summary reports. Although hard copies of these reports are sometimes available from NCES, they encourage researchers to obtain them from the CD or Web site. In addition, any results, using either SPSS, SAS, DAS, or the IPEDS software, will generate many pages of tables that will need to be printed.

If computer support is available to faculty and administrative staff in a college or university, it is always good to inform this group of new software you are using. While the DAS and ECB software come with clear installation and user instructions, computer support staff can often help with issues of disk space and optimum installation methods. They may also be able to help install or provide SPSS or SAS in the research office. In addition, programmers may be needed to chop up the IPEDS Analysis file into manageable record lengths for use in SPSS.

How to obtain data and reports: logistics

Appendix D provides the ordering addresses and information for obtaining the data and manuals and for adding your name to the NCES data set announcement mailing list. Names of NCES staff who can help expedite data requests and can answer questions are also listed.

Depending on whether the researcher is using the public or restricted-use files and depending on the data set, different steps need to be followed in order to obtain the data and reports.

IPEDS public-use files

Researchers can obtain and begin to use an IPEDS data set very easily if they know the year they want. Usually the latest available year is the most relevant. However, if one is going to merge IPEDS data onto other data sets such as NPSAS or BPS, the main IPEDS year needs to match the appropriate year of interest (such as the first year or only year the students were surveyed). The IPEDS year in the title of the CD or files indicates the year of the main files, such as enrollments. However, the financial aid and completions files always lag one year behind.

In order to obtain IPEDS, either order the CD or download the files from the Web. Then, the data can be immediately accessed using the IPEDS interface, SPSS/SAS, or spreadsheets.

NPSAS/BPS public-use files

Like IPEDS, the public-use data sets of NPSAS or BPS can be ordered on CD or downloaded from the Web. These files come with the DAS software to use it, so these data sets can be used almost immediately. Like IPEDS, knowing the latest or most appropriate year to order is the first step. Longitudinal studies such as BPS usually have different files for each year, plus cumulative files that include variables from all the available years up to the latest release, so it is usually sufficient to order the latest files.

Even if the researcher's goal is to use the restricted-use files, it is recommended that one obtain the public-use files as well, in order to conduct preliminary variable examinations and analyses and to produce accurate standard errors for the preliminary results. In addition, the DAS tables can be used for checking the accuracy of the weighted restricted-use results.

NPSAS/BPS restricted-use files

The restricted-use data files of NPSAS or BPS are available only on a CD, and NCES will send these CD's for free after a restricted data license is approved by the NCES Data Security Office. Obtaining this license takes some time and effort, but it represents little effort after you receive the license and data as long as you follow your own security plan. As mentioned above, it is recommended that the corresponding public-use data set be obtained along with the restricted-use data set in order to have another source of variable information, accurate results, and standard errors.

The major requirements for obtaining a restricted data license are:

1) Create a security plan for the data, following the guidelines in the *NCES Field Restricted User Data Procedure Manual.* The main purpose of this plan is to demonstrate to NCES that you can secure and use the data in a place that can only be accessed by you and any others included on the license. At a minimum, this means that the physical files, i.e. the CDs, must be locked in a secure place, and that the computer files must have passwords or other safeguards that would prevent others from accessing the files. There are different allowable procedures for stand-alone and networked computers that are detailed in the guidelines, but all require that any Internet or network access be blocked while the data files are accessible.
2) Everyone who works with the files must sign an affidavit stating that they will keep the data secure and protect the privacy of the respondents. There is a $250,000 fine and some jail time if you violate this agreement.
3) Agree to unannounced on-site inspections of the security arrangements.
4) Obtain the signature of a senior official of your college, such as the President. This is a further effort to secure the data in order to protect the privacy of the respondents. This person must be able to bind your college to respect and enforce the security plan in order to prevent any unauthorized access to the raw data. If you are part of a state college or university system, then you will be required to notify NCES if your state office requests access to or the use of the data for any reason, and to inform the requester of the penalties for violating the security agreement. (In the past, the State Attorney General had to sign, promising not to let the governing bodies request or use the raw data, but that is no longer necessary.)

Challenges when working with restricted-use data sets

Choosing variables: IPEDS

If one of the reasons to use the restricted-use data sets is to merge more IPEDS variables onto the files, then the IPEDS files must be examined to identify the desired variables. Reviewing the peer group literature, including the review in this report, can provide ideas for variables that are most relevant to your college or sector. However, it is important to remember that many important peer group variables are not in IPEDS, especially qualitative ones such as the mission of college.

The first place to look for IPEDS variables is the restricted-use data set you are using, in case the main variables you want to use are already on the file or are better than those available directly from IPEDS (See tables 1 and A1 for the IPEDS variables in NPSAS/BPS). For instance, the IPEDS enrollment variable already on BPS had fewer missing values then that same variable in

the IPEDS Enrollment or Analysis files. While the Analysis file includes the most popular variables from all the files, the record layouts of all five files of IPEDS should be examined in case there are extra variables that are relevant to your situation.[27] These layouts are on the CD, and can be printed out. However, many IPEDS variables are labeled only with their position on the questionnaire. Unfortunately, the questionnaires are not on the CD, and the only way to get more complete variable information is with the IPEDS CD ROM program interface software, which is only available for DOS/Windows operating systems.

Choosing variables: NPSAS/BPS

The task of reviewing the NPSAS/BPS variables is both much more formidable and easier. On the formidable side, there are many NPSAS variables and hundreds of BPS variables, and while there is a list of NPSAS variables in the NPSAS Methodology Report, there is no equivalent printed list of all the BPS variables.[28] The NCES *Statistical Analysis Reports* on NPSAS and BPS (available on the Web or from NCES) provide summary tables and descriptions of the major variables, and they should be reviewed in order to familiarize oneself with these variables, but there are many more variables not included there. These reports can also be used for checking the accuracy of your preliminary results. The BPS summary report, *Descriptive Summary of 1989-90 Beginning Postsecondary Students: 5 Years Later,* is also very helpful for determining how to analyze the issue of persistence in BPS.[29] It describes many of the persistence variables that have already been created from the periods of education and employment of the students in the waves of BPS, and it includes an essay on postsecondary persistence and attainment based on these data.

A preliminary examination of the BPS or NPSAS variables in the restricted-use file can be conducted on the public-use files using the software DAS (Data Analysis System). This task must be conducted on the computer screen of a PC. DAS displays the entire list of variables, and provides definitions and explanations of the sources of the variables and their unweighted percentage distributions on request. Variables and weights can then be selected to form the rows, columns, weights, and control variables for tables of percentages and means or for a correlation matrix for linear regression. DAS contains most, but not all of the variables found on the restricted-use file. DAS produces correctly weighted tables of groups or subgroups along with their standard errors and weighted sample sizes, which can be printed out. Thus, preliminary results of potential

---

[27] In addition, much of the Analysis file may be inaccessible to some users. It was written onto a file of 8,000 plus columns, and while the program interface software can read those columns, some versions of SPSS cannot read past 1024 columns. Therefore, programmers have to be available to chop up the file into 9 records so SPSS can read it. A suggested layout for this task is provided in appendix table A2.

[28] U.S. Department of Education, OERI, National Center for Education Statistics, Technical Report, *Methodology Report for the National Postsecondary Student Aid Study, 1992-93,* NCES 95-211, November, 1995.

[29] U.S. Department of Education, OERI, National Center for Education Statistics, Statistical Analysis Report, *Descriptive Summary of 1989-90 Beginning Postsecondary Students: 5 Years Later,* NCES 96-155, May 1996.

analyses can be checked before investing time in the restricted-use file. Actual sample sizes are not shown, but the results for any given estimate are suppressed if the sample size of that cell is less than 30.

Choosing variables from the restricted-use files is an enjoyable but intensive computer screen task given the volume of variables in BPS. The software provided with the restricted-use files, ECB (Electronic Codebook), displays a list of all the variables and weights in the files. The user can then scroll down the list and request the frequencies and percentages of the unweighted sample for each variable. The variable can then be tagged for inclusion in a codebook (variable names and value labels) and an SPSS or SAS program. While a preliminary group of variables can be selected in the first session, and then another group can be selected at another session, it is best to select all potential variables the first time, so that the different codebooks and the SPSS/SAS programs produced from each session do not need to be merged. In addition to student variables, the weights must also the chosen at this time for the SPSS/SAS program. In BPS, there are four potential weights to use, and it is best to select all of them until it can be determined which is the best to use for your purposes (see section in this chapter on choosing weights). After all potential variable and weights have been selected, the codebook can be printed, and Macintosh users can transport the SPSS program to their computer for editing and running.

Choosing and normalizing weights

The NPSAS and BPS *Methodology/Technical Reports* provide information on the sampling design and the weighting variables and design effects that resulted from that design.[30] The sampling design of NPSAS/BPS (and most NCES data sets) consisted of a stratified multistage probability sample of geographic areas, schools, and students. Weights for each student reflect this complex sampling and are necessary in order to accurately estimate the population characteristics of the postsecondary students in the sample. In addition, since BPS is a longitudinal study that includes several waves of interviews and some attrition, different weights must be used depending on whether the estimates of interest are considered cross-sectional, longitudinal, or retrospective during the different years of the study. For a clear description of these weights, see Appendix C of the BPS Statistical Analysis Report.[31] Both that report and this study used the weight, BPS94AWT, which is both cross-sectional and retrospective and is the primary weight for students who started in 1989-90 and were still alive in 1994. By using the same weight as the descriptive reports, the preliminary results from the restricted-use files could be checked against

---

[30]U.S. Department of Education, OERI, National Center for Education Statistics, Technical Report, *Beginning Postsecondary Students Longitudinal Second Follow-up (BPS:90/94) Final Technical Report*, NCES 96-153, May 1996 and U.S. Department of Education, OERI, National Center for Education Statistics, Technical Report, *Methodology Report for the National Postsecondary Student Aid Study, 1992-93*, NCES 95-211, November, 1995.

[31]U.S. Department of Education, OERI, National Center for Education Statistics, Statistical Analysis Report, *Descriptive Summary of 1989-90 Beginning Postsecondary Students: 5 Years Later*, NCES 96-155, May 1996.

these reports for accuracy. However, the DAS tables can also be used as a check for accuracy, particularly since the user can select the same weight to use in DAS that has been selected for the restricted-use analysis.

In order to produce the correct SRS standard errors for use in estimating the complex standard errors (see next section), the weights for each case must be normalized. Normalized weights apply the same proportional weight to each case, but these weights have been reduced proportionately so that the cases sum to the actual sample size rather than to the estimated population. This is necessary because the population size would produce standard errors that would be artificially low. The following formula for normalizing the weights must be applied to *each* case:

Normalized weight = (sample size / sum(weights)) x (weight of case)

The formula shows that the actual sample size is divided by the sum of the weights and then this result (which is the same for each case) is multiplied by the case's weight (which is different for each case).

Estimating standard errors

Another result of the complex sampling design of NPSAS and BPS is that standard errors of estimates, needed for any confidence intervals or tests of differences between estimates, cannot be calculated in the usual way—as if the sample was a simple random sample (SRS). Instead, advanced statistical methods that take the sample design into account, such as a Jackknife replication technique or the Taylor series method, must be used to calculate the correct standard errors. The Taylor series method is built into the DAS software in the public-use files, so any tables produced by DAS include the correct standard errors. For restricted-use data users, replicate weights are included on the restricted-use data set for use with the Jackknife technique. However, most users do not have the time or training to perform this technique, and they must find a way to estimate the correct standard errors for any analyses that cannot be duplicated by DAS.

Special software is available for this task. SUDAAN is a commercially available product developed and sold by the Research Triangle Institute (RTI) in North Carolina. WesVarPC (for Windows systems) is freeware that can be downloaded from Westat's Web site (www.westat.com). WesVarPC is used to create balanced repeated replicate weights using the replicate codes in the data file. In addition, SPSS will soon distribute (sell) a fancier version of WesVarPC as a standalone program separate from the general SPSS program.

If obtaining special software is not an option, another alternative is to use design effects with the SRS standard errors to approximate the correct standard errors.[32] For each estimate

---

[32]This technique is suggested in U.S. Department of Education, OERI, National Center for Education Statistics, *National Education Longitudinal Study (NELS) of 1988: First Followup: Student Component Data File User's*

(proportion, mean, regression coefficient), the design effect expresses the ratio of the standard error for a complex sample to the standard error for a simple random sample. Therefore, the SRS standard error can be multiplied by the design effect to produce the correct standard error for a complex sample. SRS standard errors can be obtained from the usual software or calculated in a spreadsheet.[33] However, before calculating the SRS standard errors, the weights must be normalized. See the previous section on normalizing weights. The design effects can be obtained from the published tables in the technical reports.[34] Unfortunately, design effects are provided only for the major variables and categories. In the BPS Technical report, design effects were provided only for public two-year students on the three values of persistence/attainment and the four values of highest level of education. These design effects ranged from 1.1 to 1.9.[35]

In this study, some of the student characteristics and outcomes cross tabulations (those by sector and by enrollment size) could be duplicated in DAS, and the DAS standard errors were used. For the other cross tabulations (those by diversity and percentage full-time) that could not be duplicated in DAS, the correct standard errors were estimated using the SRS standard errors and approximated design effects. Design effects were approximated by borrowing the design effects from the estimates produced in DAS for public two-year colleges by enrollment size, and applying them to the same variable for public two-year colleges by diversity and percentage full-time level. These design effects were calculated for each enrollment category by taking the ratio of the correct standard error to the SRS standard error. Then the design effect from the enrollment category with the same sample size as the diversity or full-time status category was multiplied times the SRS standard error in that category. The resulting standard errors are not perfectly accurate, because design effects differ depending upon the subgroup, the type of variables, and the range of the estimates, which are not equivalent for the different peer variables. However, they inflate the SRS standard errors enough to provide a reasonable basis for tests of difference for most institutional research purposes. Examples of calculating standard errors using design effects are provided in appendix table C1. The design effects from the DAS results ranged from .8 to 2.2, with a mean of 1.2. Since this range and mean includes the published overall design effects range of 1.1 to 1.9, this design effect mean could probably have been used for all the estimates for a rougher approximation.

Manual, Volume I, in the Sample Design chapter, section 3.6.2: Design Effects and Approximate Standard Errors, NCES 92-030, March 1992, p. 56.

[33]For proportions such as those generated in this report, the formula for the s.e. of a proportion—SQRT(pq/n)—was used. See Hubert M. Blalock, Jr., *Social Statistics* (New York: McGraw-Hill Book Company, 1972), p 195; or T.H. Wonnacott & R.J. Wonnacott, *Introductory Statistics*, Third Edition (New York: John Wiley & Sons, 1977) p 168.

[34]Design effects are also provided with the correlation matrices in the DAS output, but this is only for the variables in the public-use file, which already have their correct standard errors calculated by DAS.

[35]U.S. Department of Education, OERI, National Center for Education Statistics, Technical Report, *Beginning Postsecondary Students Longitudinal Second Follow-up (BPS:90/94) Final Technical Report*, NCES 96-153, May 1996, pp. 106-107, 112-133.

Testing for differences

Statistical tests of difference are used when a researcher wants to know whether a difference between two estimates, such as in this study between two proportions in two different types of colleges, is significant, i.e., would not occur just by chance. If the difference is significant, then that means that this level of this particular student characteristic or outcome has a high chance of being different in the two types of colleges. These tests are usually applied when the differences are substantively different, i.e. one would not test for a difference of less than six percentage points even if it might be statistically significant, because such a difference would not mean much to colleges.

One way to test for differences between proportions is to calculate the **Student's *t* statistic**, which uses the standard errors to test if the difference is significant. When the estimates are independent, such as the estimates in this study of students in different types of colleges, the ***t*** **statistic** is calculated with the formula

$$t = (p1-p2)/SQRT((se1)^2+(se2)^2)$$

and the two-tailed ***t*** distribution is used to determine significance.Since the ***t*** distribution approximates the normal curve after a sample size of 120, and all the subgroups in this study were 250 or more, the near-normal values of 1.98 and 2.62 were used to indicate the level of significance of less than .05 and .01 respectively.[36]

One more statistical adjustment needs to be made if tests for difference are performed for more than one estimate in a "family" of estimates. A family consists of all the values in each dependent variable, in this case the student characteristic/outcome variables, that are estimated in all of the categories of the independent variables, in this case institutional variables such as enrollment or diversity. The more tests that are conducted in each family, the higher the likelihood that a test will be significant just by chance. Therefore, a technique is suggested in the BPS Statistical Analysis Report that makes the test harder to attain significance if more than one test is performed in a family.[37] Starting with the acceptable level of significance of $p < .05$ for one comparison in a family, that level is divided by the number of comparisons, so that if 2 comparisons were made in the four-level SES quartiles family, then .05/2 = .025 would be the acceptable level of significance. If five comparisons were made, then the level of significance would be .05/5=.01. Since more than 5 comparisons were rarely made in any family in this study, the significance level of $p < .01$ was used for all comparisons.

---

[36]Hubert M. Blalock, Jr., *Social Statistics* (New York: McGraw-Hill Book Company, 1972) and U.S. Department of Education, OERI, National Center for Education Statistics, Statistical Analysis Report, *Descriptive Summary of 1989-90 Beginning Postsecondary Students: 5 Years Later*, NCES 96-155, May 1996, Appendix C.

[37]U.S. Department of Education, OERI, National Center for Education Statistics, Statistical Analysis Report, *Descriptive Summary of 1989-90 Beginning Postsecondary Students: 5 Years Later*, NCES 96-155, May 1996, Appendix C.

In this study, this test was applied to differences that were substantial and potentially important. Significant differences are indicated in tables 12a through 16b and the results are discussed in Chapter III. Examples of calculating ***t* statistics** are provided in appendix table C1.

Summary of steps for obtaining and working with restricted-use files

This section provides an outline of all the steps taken in this study in order to work with the restricted-use files of BPS and to merge IPEDS data onto these restricted-use files. Researchers who want to conduct some but not all of these tasks can choose the relevant sections.

Obtain relevant data sets and restricted data license

Request BPS/NPSAS public-use files

Request IPEDS public-use for corresponding year files

Apply for restricted data license and request BPS/NPSAS restricted-use files

Review variable and weighting information

IPEDS institutional variables (if adding IPEDS data to NPSAS/BPS files)

Review peer group literature

Review record layouts of the six IPEDS files to find relevant variables

Note which IPEDS variables are already in BPS

List/choose likely variables that will distinguish colleges in same sector

NPSAS/BPS Variables

Review NPSAS/BPS *Statistical Analysis Reports* to become familiar with variables

Review NPSAS and BPS *Methodology/Technical Reports* for information on sampling design, weighting variables, design effects

Use DAS in public use files to check weighted percentages of specific variables.

Use ECB in restricted-use files to examine unweighted frequencies/percentages

Select all potential variables for preliminary codebook and SPSS/SAS program

Select all potential weights for preliminary codebook and SPSS/SAS program

Print codebook and edit SPSS/SAS program for processing.

Choose/create preliminary BPS samples and files.

Run preliminary weighted frequencies and crosstabs to match DAS, summary reports

Choose IPEDS variables in BPS to indicate sample by sector or peer groups

Choose appropriate weight(s) based on summary and technical reports; normalize weight

Choose student variables of interest

Select sample by sector; create preliminary data files and new & recoded variables

Merge IPEDS variables onto BPS file

Select preliminary institutional variables from the relevant IPEDS files.

Create IPEDS files of selected variables in selected sector(s) and merge into one file.

Examine preliminary peer group indicators and groups at the institutional level.

Create school file from BPS student file for matching to IPEDS.

Match/merge IPEDS school file onto BPS school file; note extent of missing IPEDS data.

Merge BPS/IPEDS school file back onto BPS student files.

Produce final estimates

Test preliminary IPEDS peer groups with student data

Produce statistics by sector and/or peer group for student variables of interest

Revisit ECB if necessary for more student variables

Determine best way to obtain correct standard errors

Use DAS, SUDAAN, or WestVarPC

Estimate standard errors with approximate design effects

Test for differences and adjust significance level for families of estimates

Chapter V
Conclusion

V. Conclusion

Are NCES data sets *useful* to institutional researchers?

IPEDS

IPEDS can be useful to institutional researchers who want to create peer or comparison groups for their colleges for research purposes. While the peer group variables are not as detailed as the peer group literature, many institutional variables are available that can create groups of colleges that are more specific than all colleges in the same sector.

BPS/NPSAS

BPS/NPSAS can be very useful to institutional researchers who want national or peer group estimates of common student characteristics and outcomes. The BPS data set contains so many student characteristics and outcome variables that the researchers will find variables relevant to their own college. In addition, BPS and NPSAS contain major IPEDS variables, so student variables can be estimated for common institutional characteristics, such as sector and enrollment size.

Are NCES data sets *accessible* to institutional researchers?

IPEDS

IPEDS files are very accessible in that they are very easy to obtain and to use. Either a DOS/Windows or Macintosh operating system can be used to access the data. However, in order to use the NCES software to access IPEDS, a DOS/Windows system is required. IPEDS is easiest to use if one is examining institutional characteristics separate from other data sets. However, these data can be merged onto student data in other data sets, but this can be done only if one is using a restricted-use data set for the student data.

BPS/NPSAS

The public-use files of BPS/NPSAS are very easy to obtain and use, although a DOS or Windows operating system is necessary. The software provided with these files can produce estimates of the many student variables by sector or other institutional factors, including the correct standard errors and weighted sample sizes. It also provides correlation matrices that can be used to run regression analyses. However, variables cannot be recoded except for consecutive values, IPEDS or other data cannot be merged onto these data sets, and other more exploratory analyses cannot be performed.

The restricted-use files of BPS/NPSAS are hard to obtain and to use, but can be used by either DOS/Windows or Macintosh operating systems. A restricted data license must be obtained, security must be maintained, and researchers must learn more about the data set and the sampling

design and be much more involved in programming in order to produce accurate estimates. However, it is worth it to researchers who have the characteristics outlined in the next section.

Who should use the restricted-use and public-use data sets?

NCES has done an impressive job of making the national educational data sets easily accessible to researchers by providing them at three distinct levels of availability —as raw data, as customized tables, and as published (paper and electronic) tables. With these options, institutional researchers can find the level that best suits their purposes and resources.

Institutional researchers and institutional research offices with the following array of interests and resources will best enjoy working with the restricted-use data sets. First, they must have the time and resources to devote to the many programming and analysis decisions and tasks. (See the section in Chapter IV on time and other resources.) Second, they must enjoy working with raw data, creating tables tailored to their own needs, and conducting their own unique analyses. (See the section in Chapter IV on working with restricted-use data files.). This is the benefit of working with these data sets, so this is the aspect that should be enjoyed.

Researchers who want to develop their own tables geared to their own sector or type of college but do not have the time or resources for working with the restricted-use files will enjoy using the DAS software on the public-use files. DAS produces tables and standard errors for most variables or subgroups of interest, as long as the sample size does not fall below 30.

What are the alternatives?

The alternatives to working with either of the data files are to use the published reports and tables on these data sets. In this case, the researcher will be limited to the list of variables included in the report, which is much smaller than the complete list of variables. In addition, the published reports do not tend to divide up the sample farther than by sector, so within-sector subgroups cannot be found, especially for public two-year colleges.

Ironically and luckily, one of the best ways to get these reports is to order the public-use data file on a CD or to download it from the Web. Then, the report and published tables may pique researchers' interest enough to try using the DAS software to generate their own custom tables. And who knows? A successful DAS experience might encourage a foray into the world of restricted-use data sets as well.

Appendix A
Variable Lists

Table A1. Variables in IPEDS with potential for use in identifying peer groups

IPEDS variables in BPS

CTYPE	Institution type 1990 [selected only 2 & 4 yr pub, priv 4 yr]
OFCO8990	Level & control, AY89-90 ref (NPSAS)inst [lots of missing]
OFCON2	BPS Stratum - 1994 [none missing]
CTRL8990	Institution control AY89-90 [some private 2 <2 in pub 2]
SCHL8990	Institution level AY1989-90 [some 4 yr in pub 2 yr]
CALSYS	Calendar system (IPEDS)-N90
ENRLCATB	Control and size (total enrollment)-N90
FIPS	State where inst. is located (IPEDS)-N90
OBEREG	Region (OBE code) of inst. (IPEDS)-N90
OFCON1	Type and control of institution-N90
ADMREQ1	Require HS diploma/equiv. (IPEDS)-N90
ADMREQ10	Require TOEFL or equivalent (IPEDS)-N90
ADMREQ2	Require HS class standing (IPEDS)-N90
ADMREQ3	Require test scores (IPEDS)-N90
ADMREQ4	Require SAT (IPEDS)-N90
ADMREQ5	Require ACT (IPEDS)-N90
ADMREQ6	Require other test (IPEDS)-N90
ADMREQ7	Require residence (IPEDS)-N90
ADMREQ8	Require ability to benefit (IPEDS)-N90
ADMREQ9	Require age (IPEDS)-N90

IPEDS variables in IPEDS*

Enrollment total and categories
Enrollment percentage by race-ethnicity and gender
Enrollment percentage of first-time first-year students
Enrollment percentage of full-time students
Enrollment percentage of degrees awarded
Percentage of degrees awarded by race-ethnicity and gender
FTE size categories
Expenditure ratios per FTE
Program offerings by CIP
Accreditation
Fees, tuition, and charges
Categorical expenditures
Number of full-time faculty
Number faculty on tenure/non-tenure
Selected student services
Rural/urban mix
Revenue and expenditure rates
Research expenditures
Funding control
Number of campuses
Mix of high/low cost programs
Credit FTE students/FTE staff
Headcount students/FTE staff
Part-time instructional faculty & part-time staff/FTE staff

*Most of these variables are derived from several variables, so the original variable names are not listed.

Table A2. Record Layout for IPEDS 1990 Analysis File with shorter records

Original record layout

Record 1: Columns 1 to 8114

New record layout with under 1025 columns in each record

Record 1: Columns 1 to 1019
Record 2: Columns 1020 to 2037
Record 3: Columns 2038 to 3045
Record 4: Columns 3046 to 4057
Record 5: Columns 4058 to 5067
Record 6: Columns 5068 to 6043
Record 7: Columns 6044 to 7059
Record 8: Columns 7060 to 8055
Record 9: Columns 8056 to 8114

Table A3. Student characteristics and outcomes variables examined in BPS

Student Variables - BPS

Educational goal

DEGGOAL	Degree goal [includes other undergrad- fewer missing]
GOAL8990	Deg work toward, AY89-90 ref (NPSAS)inst [4 cats only]
GOAL89AB	Degree direction & inst. type in AY89-90 [same as above]
ASPIRE	Educational aspirations 1990 [highest possible - 4 cats]
PROGTYP	Degree program AY89-90-N90 [which degree it leads to]

Major/college programs

VOCA8990	Academic/vocational program AY89-90 [69% on aca]
MAJ8990	Major at AY89-90 ref (NPSAS) institution [12 cats - 37% missing]
PROGTYP	Degree program AY89-90-N90 [which degree it leads to]
MAJORS3	Major field of study AY89-90 CofEd-N90 [12 cats - 25% missing]
HBCU	Student attended hist. black inst.-N90

Type of HS/Time between high school and college

H_HSDIP	Type of H.S. diploma-B94
DELAYENR	Delayed enrollment [yes or no]
DELAYED	Type of delayed enrollment [yes is w/ & w/out hs dip]

Type and amount of college attendance

ATT8990	Intensity, AY89-90 ref (NPSAS) inst [full/parttime- 20%missing]
STAT8990	Enrollment pattern AY89-90 [detail of all ft/pt or mixed]

Experience in basic skills

REMMATH	Remedial math-N90 [seems to be a score]
REMREAD	Remedial reading-N90 [seems to be a score]
REMWRITE	Remedial writing-N90 [seems to be a score]

Level of work and financial support

ATTNXDEP	Intensity & dependency AY89-90-N90 [FT/PT with dependent]
STUIND1	Job industry AY89-90-N90 [22 cats]
STUOCC1	Job occupation AY89-90-N90 [14 cats]
STUOCC2	Job occupation AY89-90 CofEd-N90 [7 cats]
WRKHRS	Hours per week of paid work

Family obligations

KIDS8990	Num of children, AY89-90 ref (NPSAS)inst
MAR8990	Marital status, AY89-90 ref (NPSAS) inst [4 cats - 76% single]
SING8990	Single parent, AY89-90 ref (NPSAS) inst [only 6% are]
MARITALN	Marital status-N90 [3 cats]

Overall risk factors

ATRS8990	# risk factors, AY89-90 ref (NPSAS) inst [0-7]
RISKNDX2	Sum of 7 risk factors-comp to NP87/90 [0-7 - same?]

Family SES, educational and occupational background

RFATHED	Father's education recoded [7 cats for each]
RMOTHED	Mother's education recoded
RPARED	Parent's education recoded (maximum)
DADOC1	Father's occupation (recoded)-N90 [7 cats]
MOMOC1	Mother's occupation(recoded)-N90 [7 cats]
SESPERC	Socioeconomic status (SES) percentile-N90
SESGRP	SES in 4 groups

Other student demographic characteristics:

Gender

H_GENDR	Gender-B94

Race-ethnicity

BPSRACE	Derived race/ethnicity from BPS [5 cats w/ hispanic]
BRACESEX	Race-ethnicity and gender of student.
H_RACE	Race-B94 [FIVE CATS- no hispanic]
H_ASIAN	Asian type-B94
H_HISP	Whether Hispanic-B94
H_HISPT	Hispanic type-B94

Citizenship

H_CITIZ	Whether U.S. citizen-B94

Age

AGE	Age as of Dec 31, 1989

Outcomes

EI_GRYR1	Grades for AY89-90-B92

First year retention/persistence variables

PER8990R	Persist and attain AY89-90 (re PERA8990) [result at end of yr]
PERA8990	Persist and attain - AY89-90 [mix of 1st yr and after 5 yrs]
PERADEG	Degree progression [Degree status of all degrees after 5 yrs]
PAABA90	Persisted or transfered from AA to BA after 89-90x
PAABA91	Persisted or transfered from AA to BA after 90-91
PAABA92	Persisted or transfered from AA to BA after 91-92
PAABA93	Persisted or transfered from AA to BA after 92-93

Appendix B

Supporting tables for Chapter II

Table B1. Standard errors for percentage distribution of race-ethnicity of 1989-90 beginning postsecondary students, by selected sectors

	Public 4-year	Private 4-year	Public 2-year
White	1.64	1.33	2.29
African American	1.26	1.05	1.30
Latino	.74	.52	1.49
Asian/Pacific Islander	.76	.52	.88
Native American	.16	.22	.32

SOURCE: U.S. Dept. of Education, OERI, NCES, Beginning Postsecondary Students Longitudinal Study (BPS: 90/94), Public-Use File, Data Analysis System (DAS)

Table B2. Standard errors for percentage distribution of race-ethnicity of 1989-90 beginning postsecondary students in public 2-year colleges, by enrollment size, selected regions, and selected states

	Enrollment size				
	Under 2,500	2,500-4,999	5,000-9,999	10,000-19,999	20,000 plus
White	2.64	5.27	4.65	5.54	9.42
African American	2.05	4.88	2.51	1.48	4.85
Latino	1.46	1.98	3.40	3.48	7.40
Asian/Pacific Islander	.00	.83	1.36	2.28	5.10
Native American	.91	.00	.67	.77	.00

	Selected regions			
	Middle East	Great Lakes	South East	Far West
White	5.25	4.00	2.91	4.51
African American	4.46	3.30	2.56	2.50
Latino	2.30	2.01	1.41	3.48
Asian/Pacific Islander	.93	1.39	.71	2.46
Native American	.00	.91	.39	1.12

	Selected states			
	New York	Texas	Florida	California
White	8.33	6.65	3.58	4.86
African American	4.53	5.17	3.48	2.86
Latino	5.42	6.40	2.77	3.78
Asian/Pacific Islander	1.20	4.87	1.05	2.63
Native American	.00	.00	.00	1.18

SOURCE: U.S. Dept. of Education, OERI, NCES, Beginning Postsecondary Students Longitudinal Study (BPS: 90/94), Public-Use File, Data Analysis System (DAS)

Table B3. Percentage distribution of public two-year colleges in IPEDS 1990, by selected student demographic and outcome measures

Percentage of women out of total enrollment	
0-4%	<1
5-19%	1
20-35%	1
36-65%	65
66-100%	13
Missing	21
Total	100%
Number of AA/AS degrees	
0-99	14
100-249	27
250-499	20
500-749	9
750-plus	7
Missing	23
Total	100%
Percentage of AA/AS degrees out of total enrollment	
0-4%	16
5-9%	38
10-14%	16
15-19%	4
20-100%	2
Missing	24
Total	100%
Percentage of women out of total AA/AS degrees	
0-39%	5
40-49%	5
50-59%	22
60-69%	31
70-79%	12
80% plus	3
Missing	23
Total	100%
Percentage of white students out of total AA/AS degrees	
0-64%	11
65-79%	11
80-89%	21
90-94%	14
95-100%	19
Missing	26
Total	100%

SOURCE: U.S. Dept. of Education, OERI, NCES, Integrated Post-secondary Education Data System (IPEDS), 1990

Table B4. Crosstabulations and chi-squared values of potential peer group variables for public two-year colleges in IPEDS 1990

October 1990 Enrollment size	Percentage of full-time students: Under 35%	Percentage of full-time students: 35% or more	Percentage of first-time students: Under 35%	Percentage of first-time students: 35% or more
Under 2,500	21	59	40	60
2,500-9,999	53	37	45	38
10,000 or more	26	4	15	2
Total	100%	100%	100%	100%
	χ^2=177.7; d.f.=2; p< .00001		χ^2=31.9; d.f.=2; p<.00001	

Percentage of first-time students	Percentage of full-time students: Under 35%	Percentage of full-time students: 35% or more
Under 35%	93	77
35% or over	7	23
Total	100%	100%
	χ^2=48.1; d.f.=1; p< .00001	

Percentage of students of color	October 1990 enrollment size: Under 2,500	2,500 to 9,999	10,000 or more
Under 20%	66	66	46
20-34%	18	19	27
35% or more	16	15	27
Total	100%	100%	100%
	χ^2=19.1; d.f.=4; p< .001		

Percentage of students of color	Percentage of full-time students: Under 35%	Percentage of full-time students: 35% or more	Percentage of first-time students: Under 35%	Percentage of first-time students: 35% or more
Under 20%	55	69	63	67
20-34%	25	16	20	15
35% or more	21	16	17	18
Total	100%	100%	100%	100%
	χ^2=17.4; d.f.=2; p< .001		χ^2=1.94; d.f.=2; p=.38	

SOURCE: U.S. Dept. of Education, OERI, NCES, Integrated Post-secondary Education Data System (IPEDS), 1990

Appendix C

Supporting tables for Chapter III

Table C1. Examples of using DAS standard errors to estimate the DAS design effect and using this design effect to estimate the standard errors for non-DAS variables and perform *t*-tests on the differences between the percentages of selected student characteristics and outcomes of beginning postsecondary students in public two-year colleges

	Enrollment size					Level of diversity				
	Wtd. nrmlzd. n	Pct. p	SRS s.e.	DAS s.e.	Design effect	Wtd. nrmzld. n	Pct. p	SRS s.e.	Est'd s.e.	
Enrollment size under 10,000	570					**Diversity less than 20%**	534			
Race-ethnicity										
White		80	1.68	3.62	2.16		89	1.35	2.92	
African American		9	1.20	2.36	1.97		5	0.94	1.86	
Latino		8	1.14	2.37	2.09		4	0.85	1.77	
Asian/Pacific Isl		2	0.59	0.93	1.59		1	0.43	0.68	
Native American		1	0.42	0.44	1.06		<1	0.27	0.29	
Enrollment size 10,000 or more	269					**Diversity 20% or more**	278			
Race-ethnicity										
White		68	2.84	4.86	1.71		51	3.00	5.12	
African American		6	1.45	1.54	1.06		15	2.14	2.28	
Latino		16	2.24	3.17	1.42		25	2.60	3.68	
Asian/Pacific Isl		9	1.74	2.09	1.20		9	1.72	2.06	
Native American		1	0.61	0.61	1.01		1	0.60	0.60	

Student *t* statistics for percentages in colleges with low vs. high diversity

	Pct. low	Pct. high	s.e. low	s.e. high	*t* statistic	Level of significance
Table 12a						
Pct. white: diversity low/high	89	51	2.92	5.12	6.45	$p < .01$
Pct. Afric Am: diversity low/high	5	15	1.86	2.28	-3.40	$p < .01$
Pct. Latino: diversity low/high	4	25	1.77	3.68	-5.14	$p < .01$
Pct. Asian: diversity low/high	1	9	0.68	2.06	-3.69	$p < .01$

NOTES: Wtd. nrmlzd. n: Actual sample size of peer variable category. Weights have been reduced proportionately and are still in effect.
Pct p: The weighted proportion of students in this category within this peer variable category
SRS s.e.: Standard error calculated as if sample was simple random sample (SRS)
DAS s.e.: Standard error calculated by DAS to take into account complex sample design
Design effect: DAS s.e. divided by SRS s.e.
Est'd s.e.: SRS s.e. multiplied by design effect from DAS variable with similar sample size and p
For more explanation of the elements in this table, see the sections on estimating standard errors and testing for differences in Chapter IV.

SOURCE: U.S. Dept. of Education, OERI, NCES, Beginning Postsecondary Students Longitudinal Study (BPS: 90/94), Public-use File and Restricted-use File; U.S. Dept. of Education, OERI, NCES, Integrated Post-secondary Education Data System (IPEDS), 1990

Table C12a. Standard errors for percentage distribution of gender, race-ethnicity, citizenship, age, and religion of beginning postsecondary students, by sector and by percentage of students of color within public two-year colleges, 1989-90

	Sector			Public two-yr colleges Percentage of students of color	
	Public 4-yr	Private 4-yr	Public 2-yr	Under 20%	20% or more
Sample size:	1,889	2,499	889	534	278
	Calculated by DAS			Estimated with des. effects*	
Gender					
Male	1.24	1.53	1.93	2.62	3.30
Female	1.24	1.53	1.93	2.62	3.30
Race-ethnicity					
White, non-Latino	1.64	1.33	2.29	2.92	5.12
African-Am, non-Latino	1.26	1.05	1.30	1.86	2.28
Latino	.74	.52	1.49	1.77	3.68
Asian/Pacific Islander	.76	.52	.88	.68	2.06
Native American	.16	.22	.32	.29	.60
U.S. Citizenship					
U.S. Citizen	.43	.35	.68	.59	1.76
Not a U.S. Citizen	.43	.35	.68	.59	1.76
Age					
19 yrs or less	.83	1.01	2.29	2.95	3.35
20-24 yrs	.61	.59	1.37	1.53	2.54
25-29 yrs	.25	.29	1.08	1.32	2.15
30-39 yrs	.36	.37	.94	1.33	1.41
40 yrs or more	.21	.29	1.00	1.37	1.71
Religion					
Protestant	1.75	1.97	2.17	1.42	2.25
Catholic	1.57	2.55	2.20	2.75	3.65
Other religion	1.03	1.98	1.68	2.08	2.95
None	.70	.75	1.39	1.77	2.76

NOTE: *For details on this procedure, see section on estimating standard errors in Chapter IV.
SOURCE: U.S. Dept. of Education, OERI, NCES, Beginning Postsecondary Students Longitudinal Study (BPS: 90/94), Public-use File and Restricted-use File

Table C12b. Standard errors for percentage distribution of gender, race-ethnicity, citizenship, age, and religion of beginning postsecondary students in public two-year colleges, by enrollment size and percentage of full-time students: 1989-90

	Enrollment size				Percentage of full-time students	
	Under 2,500	2,500 to 9,999	10,000 or more	Under 10,000	Under 35%	35% or more
Sample size:	108	402	269	510	454	351
		Calculated by DAS			Est'd with des. effects*	
Gender						
Male	4.68	3.01	3.34			
Female	4.68	3.01	3.34			
Race-ethnicity						
White, non-Latino	2.64	3.62	4.86	2.68		
African-Am, non-Latino	2.05	2.36	1.54	1.77		
Latino	1.46	2.37	3.17	1.75		
Asian/Pacific Islander	.00	.93	2.09	.66		
Native American	.91	.44	.61	.41		
U.S. Citizenship						
U.S. Citizen	1.46	.96	1.53			
Not a U.S. Citizen	1.46	.96	1.53			
Age						
19 yrs or less	4.59	3.47	3.24			
20-24 yrs	2.62	1.83	2.54			
25-29 yrs	2.76	1.62	1.57			
30-39 yrs	1.72	1.58	1.22			
40 yrs or more	2.05	1.67	1.25			
Religion						
Protestant	4.96	3.17	3.63			
Catholic	4.26	3.22	3.78			
Other religion	3.93	2.39	3.02			
None	1.90	2.17	2.87			

NOTE: *For details on this procedure, see section on estimating standard errors in Chapter IV.
SOURCE: U.S. Dept. of Education, OERI, NCES, Beginning Postsecondary Students Longitudinal Study (BPS: 90/94), Public-use File and Restricted-use File

Table C13a. Standard errors for percentage distribution of SES, parent's education, high school diploma and years since high school, marital status and number of children of beginning postsecondary students, by sector and by percentage of students of color within public two-year colleges, 1989-90

	Sector			Public two-yr colleges Percentage of students of color	
	Public 4-yr	Private 4-yr	Public 2-yr	Under 20%	20% or more
Sample size:	1,889	2,499	889	534	278
	Calculated by DAS			Estimated with des. effects*	
Socio-economic status (SES)					
Lowest quartile	.75	.65	1.59	1.98	2.61
Low middle quartile	.76	.73	1.65	1.85	3.09
High middle quartile	1.20	1.07	1.64	2.17	2.74
Highest quartile	1.52	1.66	1.95	2.42	3.08
Parent's highest education					
High school or less	1.36	1.52	2.02	2.51	3.38
Trade school	.61	.49	.73	.89	.92
Some college	1.12	1.07	1.47	1.81	2.96
Bachelor's degree	1.25	1.01	1.44	2.08	2.66
Graduate/prof. degree	1.12	1.77	1.28	1.59	2.50
Type of high school diploma					
Regular diploma	.32	.34	1.09		
Other or none	.32	.34	1.09		
Number of years since high school graduation					
Same year	.94	.97	2.24		
1-2 years	.72	.49	1.25		
3-9 years	.41	.44	1.26		
10 years or more	.44	.54	1.51		
Marital status					
Not married	.48	.63	1.80		
Married	.47	.61	1.75		
Separated	.12	.10	.48		
Number of children					
None	.52	.61	1.68		
One	.35	.29	1.09		
Two or more	.35	.44	1.29		
Single parent					
Yes	.27	.34	.88		
No	.27	.34	.88		

NOTE: *For details on this procedure, see section on estimating standard errors in Chapter IV.
SOURCE: U.S. Dept. of Education, OERI, NCES, Beginning Postsecondary Students Longitudinal Study (BPS: 90/94), Public-use File and Restricted-use File

Table C13b. Standard errors for percentage distribution of SES, parent's education, high school diploma and years since high school, marital status and number of children of beginning postsecondary students in public two-year colleges, by enrollment size and percentage of full-time students: 1989-90

	Enrollment size				Percentage of full-time students	
	Under 2,500	2,500 to 9,999	10,000 or more	Under 10,000	Under 35%	35% or more
Sample size:	108	402	269	510	454	351
	Calculated by DAS				Est'd with des. effects*	
Socio-economic status (SES)						
Lowest quartile	3.10	2.37	2.33			
Low middle quartile	3.87	2.31	2.79			
High middle quartile	3.42	2.46	2.84			
Highest quartile	4.49	2.66	3.45			
Parent's highest education						
High school or less	4.47	2.93	3.31	2.47		
Trade school	1.96	1.04	1.10	.95		
Some college	2.90	2.11	2.78	1.71		
Bachelor's degree	2.53	2.37	2.55	1.82		
Graduate/prof. degree	1.52	2.03	2.45	1.54		
Type of high school diploma						
Regular diploma	1.57	1.52	2.06	1.17		
Other or none	1.57	1.52	2.06	1.17		
Number of years since high school graduation						
Same year	5.15	3.10	3.28	2.74		
1-2 years	2.00	1.84	2.45	1.50		
3-9 years	3.23	1.69	2.15	1.53		
10 years or more	2.93	2.37	2.09	1.91		
Marital status						
Not married	4.37	2.63	2.10	2.26		
Married	4.20	2.57	1.99	2.20		
Separated	1.46	.52	.98	.58		
Number of children						
None	3.88	2.50	2.41	2.10		
One	2.60	1.73	1.67	1.44		
Two or more	3.14	1.89	1.86	1.62		
Single parent						
Yes	2.17	1.40	1.42	1.17		
No	2.17	1.40	1.42	1.17		

NOTE: *For details on this procedure, see section on estimating standard errors in Chapter IV.
SOURCE: U.S. Dept. of Education, OERI, NCES, Beginning Postsecondary Students Longitudinal Study (BPS: 90/94), Public-use File and Restricted-use File

Table C14a. Standard errors for percentage distribution of distance from home, residence, financial dependency, college attendance, and major of beginning postsecondary students, by sector and by percentage of students of color within public two-year colleges, 1989-90

	Sector			Public two-yr colleges Percentage of students of color	
	Public 4-yr	Private 4-yr	Public 2-yr	Under 20%	20% or more
Sample size:	1,889	2,499	889	534	278
	Calculated by DAS			Estimated with des. effects*	
Distance of college from home					
5 miles or less	.93	.76	1.92	2.48	3.28
6-10 miles	.91	.61	1.87	2.36	3.58
11-50 miles	1.87	1.64	2.05	2.79	2.95
51-100 miles	1.15	.97	.62		
101-500 miles	2.04	1.55	.94		
Over 500 miles	.69	1.62	.14		
Type of residence					
Campus housing	2.43	1.96	.99		
Off-campus	1.05	.98	2.32		
With parents	2.15	1.47	2.31		
Financial dependency status					
Dependent	.86	.96	2.10		
Independent	.86	.96	2.10		
College attendance status					
Full-time	1.39	.99	2.42	3.00	3.37
Half-time	1.04	.74	1.82	2.03	2.93
Less than half-time	.61	.50	2.02	2.14	3.53
Major in 1989-90					
Humanities	.90	1.37	1.68		
Social and behav. sci.	1.00	1.45	.88		
Life sciences	.64	.64	.80		
Physical sciences	.36	.45	.16		
Mathematics	.36	.22	.25		
Computer and info sci	.06	10	.33		
Engineering	1.01	1.47	1.51		
Education	.96	.88	1.07		
Business & mgmnt.	1.12	1.49	1.94		
Health	.73	.93	1.41		
Vocational/tech.	.49	.37	1.35		
Other technical	.84	.62	1.32		

NOTE: *For details on this procedure, see section on estimating standard errors in Chapter IV.
SOURCE: U.S. Dept. of Education, OERI, NCES, Beginning Postsecondary Students Longitudinal Study (BPS: 90/94), Public-use File and Restricted-use File

Table C14b. Standard errors for percentage distribution of distance from home, residence, financial dependency, college attendance, and major of beginning postsecondary students in public two-year colleges, by enrollment size and percentage of full-time students: 1989-90

	Enrollment size				Percentage of full-time students	
	Under 2,500	2,500 to 9,999	10,000 or more	Under 10,000	Under 35%	35% or more
Sample size:	108	402	269	510	454	351
	Calculated by DAS				Est'd with des. effects*	
Distance of college from home						
5 miles or less	3.75	3.05	3.21		2.83	3.19
6-10 miles	4.46	2.78	3.56		2.88	2.69
11-50 miles	4.73	3.21	3.09		2.98	3.45
51-100 miles	2.15	.81	.73			
101-500 miles	4.23	.82	.80			
Over 500 miles	.00	.31	.00			
Type of residence						
Campus housing	4.50	.75	.41			
Off-campus	5.21	3.60	3.41			
With parents	4.87	3.59	3.42			
Financial dependency status						
Dependent	4.08	3.35	3.02			
Independent	4.08	3.35	3.02			
College attendance status						
Full-time	5.23	3.42	3.49		3.25	3.63
Half-time	4.32	2.35	3.08		2.43	2.30
Less than half-time	3.02	2.76	3.21		2.56	2.54
Major in 1989-90						
Humanities	3.63	2.32	3.44	1.97		
Social and behav. sci.	1.84	1.14	1.98	.97		
Life sciences	1.12	1.43	1.31	1.08		
Physical sciences	.88	.00	.00	.25		
Mathematics	.75	.00	.70	.21		
Computer and info sci	.00	.51	.78	.36		
Engineering	3.21	2.49	2.05	2.01		
Education	1.84	1.72	1.87	1.34		
Business & mgmnt.	3.81	2.82	3.91	2.33		
Health	3.86	1.86	2.63	1.73		
Vocational/tech.	2.83	1.97	2.46	1.63		
Other technical	2.45	2.20	1.85	1.72		

NOTE: *For details on this procedure, see section on estimating standard errors in Chapter IV.
SOURCE: U.S. Dept. of Education, OERI, NCES, Beginning Postsecondary Students Longitudinal Study (BPS: 90/94), Public-use File and Restricted-use File

Table C15a. Standard errors for percentage distribution of work hours, current occupation, and future occupation goal of beginning postsecondary students, by sector and by percentage of students of color within public two-year colleges, 1989-90

	Sector			Public two-yr colleges Percentage of students of color	
	Public 4-yr	Private 4-yr	Public 2-yr	Under 20%	20% or more
Sample size:	1,889	2,499	889	534	278
	Calculated by DAS			Estimated with des. effects*	
Number of hours worked for pay in 1989-90					
None	1.01	1.16	1.36		
1-19 hours	.96	1.07	1.29		
20-34 hours	1.27	.97	1.61		
35 or more hours	1.17	1.22	2.03		
Current occupation					
Professional	NOT AVAILABLE IN PUBLIC-USE FILES				
Executive					
Marketing					
Administrative support					
Technical					
Service					
Blue collar					
Occupation expected in future					
Marketing	.69	.59	.90		
Administrative support	.51	.63	1.18		
Service	.50	.48	1.10		
Executive	1.10	1.24	1.85		
Postsecondary teacher	.21	.36	.16		
Other education	1.04	1.07	1.24		
Engineer/Architect	1.07	1.28	1.25		
Scientist	.53	.61	.58		
Computer science	.35	.31	.76		
Social/recreation	.39	.60	.60		
Doctor/dentist	.51	.43	.50		
Other medical	.94	1.02	1.33		
Technical	.49	.34	1.06		
Lawyer	.54	.66	.61		
Blue collar	.60	.76	.93		
Writer/artist	.90	1.13	1.12		

NOTE: *For details on this procedure, see section on estimating standard errors in Chapter IV.
SOURCE: U.S. Dept. of Education, OERI, NCES, Beginning Postsecondary Students Longitudinal Study (BPS: 90/94), Public-use File and Restricted-use File

Table C15b. Standard errors for percentage distribution of work hours, current occupation, and future occupation goal of beginning postsecondary students in public two-year colleges, by enrollment size and percentage of full-time students: 1989-90

	Enrollment size				Percentage of full-time students	
	Under 2,500	2,500 to 9,999	10,000 or more	Under 10,000	Under 35%	35% or more
Sample size:	108	402	269	510	454	351
	Calculated by DAS				Est'd with des. effects*	
Number of hours worked for pay in 1989-90						
None	3.38	1.78	2.59	1.61	1.63	1.95
1-19 hours	2.98	1.93	2.26	1.65	1.81	2.20
20-34 hours	3.84	1.98	3.42	1.84	1.93	2.09
35 or more hours	4.24	2.78	4.20	2.40	2.59	2.97
Current occupation						
Professional	NOT AVAILABLE IN PUBLIC-USE FILES					
Executive						
Marketing						
Administrative support						
Technical						
Service						
Blue collar						
Occupation expected in future						
Marketing	1.90	1.14	1.60	.99		
Administrative support	2.84	1.76	2.00	1.50		
Service	1.90	1.75	1.79	1.42		
Executive	3.94	2.84	3.33	2.32		
Postsecondary teacher	.00	.34	.00	.24		
Other education	2.43	1.93	2.23	1.56		
Engineer/Architect	2.85	1.64	2.58	1.44		
Scientist	1.41	.74	1.31	.67		
Computer science	1.91	1.23	.90	1.03		
Social/recreation	1.47	.60	1.36	.60		
Doctor/dentist	.00	.74	.61	.53		
Other medical	3.06	1.84	2.77	1.58		
Technical	2.52	1.62	1.87	1.37		
Lawyer	1.20	.72	1.50	.63		
Blue collar	1.91	1.53	1.22	1.24		
Writer/artist	1.65	1.80	2.35	1.37		

NOTE: *For details on this procedure, see section on estimating standard errors in Chapter IV.
SOURCE: U.S. Dept. of Education, OERI, NCES, Beginning Postsecondary Students Longitudinal Study (BPS: 90/94), Public-use File and Restricted-use File

Table C16a. Standard errors for percentage distribution of degree goal, educational aspirations, persistence and attainment, and degree progression of beginning postsecondary students, by sector and by percentage of students of color within public two-year colleges, 1989-90

	Sector			Public two-yr colleges Percentage of students of color	
	Public 4-yr	Private 4-yr	Public 2-yr	Under 20%	20% or more
Sample size:	1,889	2,499	889	534	278
	Calculated by DAS			**Estimated with des. effects***	
Degree currently working toward					
None	.35	.28	1.01		
Vocational certificate	.29	.30	1.64		
Associate's degree	.93	1.05	2.26	2.60	4.41
Bachelor's degree	.99	1.19	2.11	2.24	3.93
Educational aspirations					
Trade school	.38	.39	1.63	1.84	2.51
Associate's degree	.60	.47	1.59	2.10	2.86
Bachelor's degree	1.38	1.38	1.98	2.59	3.22
Advanced degree	1.63	1.73	1.96	2.26	3.75
Persistence and attainment in first year					
Attained Certificate					
Persister to next year	1.36	1.50	2.05		
Transferred during year	.51	.51	.75		
Subsequent transfer	.88	1.02	1.44		
Stopout in 89/90	.53	.47	1.35	2.17	2.85
Left without returning	.75	.68	1.73	3.99	3.04
Degree progression in 4 years					
Attained BA	1.70	2.07	.65		
Attained AA, then BA	.25	.37	.79		
Attained AA, enrld BA	.21	.13	.62		
Attained AA, not enrld	.52	.40	1.49		
Attained certificate, then degree or enrolled	.21	.20	.36		
Attained cert, not enrlled	.46	.39	1.31		
No degree, enrolled BA	1.05	.66	.99		
No degree, enrolled AA	.32	.20	.99		
No degree, enrolled	.10	.10	.45		
No degree, not enrolled	1.44	1.40	2.06	2.43	3.97

NOTE: *For details on this procedure, see section on estimating standard errors in Chapter IV.
SOURCE: U.S. Dept. of Education, OERI, NCES, Beginning Postsecondary Students Longitudinal Study (BPS: 90/94), Public-use File and Restricted-use File

Table C16b. Standard errors for percentage distribution of degree goal, educational aspirations, persistence and attainment, and degree progression of beginning postsecondary students in public two-year colleges, by enrollment size and percentage of full-time students: 1989-90

	Enrollment size				Percentage of full-time students	
	Under 2,500	2,500 to 9,999	10,000 or more	Under 10,000	Under 35%	35% or more
Sample size:	108	402	269	510	454	351
	Calculated by DAS				Est'd with des. effects*	
Degree currently working toward						
None	1.92	1.65	1.55	1.32		
Vocational certificate	3.22	2.68	1.99	2.13		
Associate's degree	4.54	3.03	4.54	2.55		
Bachelor's degree	4.59	2.31	4.21	2.11		
Educational aspirations						
Trade school	3.68	2.20	1.83	1.89		
Associate's degree	3.14	2.53	2.65	2.02		
Bachelor's degree	4.28	3.01	3.31	2.48		
Advanced degree	2.64	2.94	3.86	2.23		
Persistence and attainment in first year						
Attained Certificate						
Persister to next year	4.53	2.78	3.64	2.44		
Transferred during year	1.82	1.15	1.24	.98		
Subsequent transfer	3.21	2.04	2.38	1.75		
Stopout in 89/90	1.62	2.05	2.66	1.58		
Left without returning	4.03	2.36	3.12	2.04		
Degree progression in 4 years						
Attained BA	2.09	.66	1.33	.78		
Attained AA, then BA	2.40	.97	1.52	.98		
Attained AA, enrld BA	1.29	1.05	.97	.83		
Attained AA, not enrld	2.28	2.06	3.29	1.62		
Attained certificate, then degree or enrolled	.83	.70	.00	.56		
Attained cert, not enrlled	2.95	1.90	2.30	1.61		
No degree, enrolled BA	1.86	1.44	2.06	1.16		
No degree, enrolled AA	1.05	1.61	1.86	1.21		
No degree, enrolled	.31	.89	.51	.65		
No degree, not enrolled	3.74	2.97	4.06	2.38		

NOTE: *For details on this procedure, see section on estimating standard errors in Chapter IV.
SOURCE: U.S. Dept. of Education, OERI, NCES, Beginning Postsecondary Students Longitudinal Study (BPS: 90/94), Public-use File and Restricted-use File

Appendix D

NCES Ordering Information

Appendix D
Contact names and addresses at National Center for Education Statistics (NCES)

To obtain Public-use Data Files

All **public use NCES data files** and **Department of Education Publications** can be found via the World Wide Web by starting at:

http://www.ed.gov/NCES/ For **data files,** follow the paths to data and surveys.
For **publications,** follow the paths to publications.

The **DAS (Data Analysis System)** software with the **NCES data files** and **codebooks** can be accessed and downloaded via the Web directly from the following address:

http://www.PEDAR-DAS.org/

To order **data**, **codebooks**, and **DAS** on **CD-ROMs**, or for special **runs of NCES data**, contact:

National Education Data Resource Center (NEDRC)
(703) 845-3151 (voice) - (703) 820-7465 (fax)
ndrc@inet.ed.gov NOTE: cc to Aurora (see below) who will make sure you get it.

To order a **single hard copy of current reports** free of charge, order from:

National Library of Education (NLE) at (800) 424-1616 **or** (202) 219-1651
LIBRARY-NLE@ed.gov **or** library@inet.ed.gov

Actual person: To E-mail or call a very helpful, friendly NCES person who takes her public information job seriously and will answer any question about obtaining or using these publications, datasets, or direct you to the person in NCES who can help, contact:

Aurora_D'Amico Aurora_D'Amico@ed.gov or adamico@inet.ed.gov
NCES Longitudinal Studies Branch
555 New Jersey Ave, NW
Washington, DC 20208-5652
(202) 219-1365

Other helpful NCES staff with specialties:

Andrew_Malizio@ed.gov (BPS and merging with IPEDS)
Dennis_Carroll@ed.gov (BPS and merging with IPEDS)
Bill_Freund@ed.gov (IPEDS)

To obtain Restricted-use Data Files

Obtain Restricted Data License (or amend current license to add data set) from:

Samuel Peng
Data Security Officer
Department of Education /NCES/SSSG
555 New Jersey Avenue, NW, Room 418
Washington, DC 20208-5654

Samuel_Peng@ed.gov
(202) 219-1643
Cynthia_Barton@ed.gov
(202) 219-2199
Ruth_Harris@ed.gov
(202) 219-1831

First, request from Cynthia or Ruth: *NCES Field Restricted User Data Procedure Manual*
Write request letter and security plan according to specifications in manual.
Call Cynthia Barton for overall questions; call Samuel Peng for any sector or site-specific questions
Time to allow from first request of manual to receipt of data on CD: About 1 month

To get on the OERI/NCES mailing list for data set release announcements

E-mail your name and address to: irene_fernandez@ed.gov **or** torey_evans@ed.gov
or call Irene Fernandez at: (202) 219-1570